Evaluating the suitability of Narrowband-Internet of Things (NB-IoT) for smart grids

by

Varun Nair

in partial fulfilment of the requirements for the degree of

Master of Science
in Electrical Engineering
Telecommunications and Sensing Systems

at the Delft University of Technology,
to be defended publicly on Thursday, November 30th, 2017 at 2 PM

Thesis committee:	Dr. Remco Litjens MSc	TU Delft, TNO
	Dr. Haibin Zhang	TNO
	Dr. Ir. Eric Smeitink	TU Delft, KPN

TU Delft NAS Network Architectures & Services Group

TNO innovation for life

Preface

This master thesis marks the culmination of my studies at the Delft University of Technology for obtaining a Master of Science (MSc) degree in Electrical Engineering (Telecommunications & Sensing Systems). The experience of working on this thesis was both challenging and intellectually stimulating. The success of this project would have been impossible without the immense support and timely guidance received from several people. I would like to, firstly, express my gratitude towards the Networks department of TNO for hosting me and providing the necessary resources to conduct my research. I had a really enjoyable time in the past one year and three months at TNO, which was also filled with great learning experiences. Needless to say, working in the company of several accomplished colleagues provided me with valuable insight about the organisation and on the state-of-art research in the field of networks.

I would like to thank Haibin Zhang, my daily supervisor at TNO, for selecting me to work on this topic at TNO, devoting time out of his schedule in providing useful feedback on my work (including writing and presentation), addressing technical queries and engaging in interesting and sometimes long discussions that helped to improve the quality of the work. My gratitude also goes to Remco Litjens, my supervisor at TU Delft, for his inputs in shaping the research goals and the approach, patience during extended phone calls/meetings and importantly, the critical feedback provided throughout the project. His detailed review on the draft chapters of the thesis provided a useful perspective of my work and greatly improved my skills in academic writing. I would like to also thank Eric Smeitink for agreeing to be part of the thesis committee. I wish to acknowledge the support received from Ljupco Jorguseski and Yohan Toh, who were kind enough to provide their inputs, whenever I needed, in the initial design of the NB-IoT simulator for the thesis. Many thanks also go to Sylvie Dijkstra-Soudarissanane for getting me acquainted with the High Performance Cluster (HPC) at TNO.

Finally, a heartfelt thanks to my family, particularly my parents back in India, and friends who gave me all their support and encouragement during the several ups and downs I went through.

Varun Nair

Abstract

Narrowband Internet-of-Things (NB-IoT), recently introduced by 3GPP, is a relevant Radio Access Technology (RAT) solution for deployment within smart grids, the electricity grids of the future, due to the need to provide low-cost connectivity to a large number of smart meters installed in households. Outage Restoration and Management (ORM), a smart grid use case, involves the smart meter User Equipment (UE) sending a notification message to the utility upon the detection of a loss or restoration of power. ORM is an effective way for utilities to quickly detect, localise and restore a power outage. However, depending on the extent of the power outage, a near simultaneous network access by multiple UEs may occur, leading to resource congestion, particularly of the so-called 'random access channel'. This may impact the reliability, i.e. the percentage of total notifications successfully delivered within a certain transfer delay target, and in turn, the accuracy of the power outage localisation. Consequently, the maximisation of the ORM reliability performance for a technology like NB-IoT becomes a challenge, given that such use cases, though relevant, were not considered in its design phase.

The main goal of this thesis is the optimisation of the NB-IoT network configuration, with a focus on packet scheduling, in order to maximise the ORM reliability performance. To this extent, a system-level simulation model is developed and implemented, incorporating realistic characteristics of energy distribution and mobile networks in four different environments (rural, suburban, urban and dense urban), the traffic characteristics of ORM and the relevant 3GPP specifications of NB-IoT. Additionally, a set of candidate time-frequency domain packet schedulers are proposed. A sensitivity analysis of key network configuration components is performed for a set of power outage scenarios i.e. network loads, the associated optimisation trade-offs are highlighted and a near-optimal network configuration is derived.

Based on the sensitivity analysis, a proposed scheduler which prioritises UEs based on a combination of the Earliest Due Date First (EDDF) and Shortest Processing Time First (SPTF) principles, and assigns each UE a single uplink subcarrier with a subcarrier spacing of 3.75 kHz, performs best amongst all the candidate schedulers. Furthermore, the achieved reliability performance is close to 100% for all the considered power outage scenarios in the rural and dense urban environments. In the suburban and urban environments, close to 100% reliability is achieved for the majority of the power outage scenarios.

List of Figures

List of Tables

Contents

Chapter 1. Introduction

In this chapter, the reader is introduced to some background, motivation and key research goals for the work presented in this thesis.

In Section 1.1, the idea of massive Internet-of-Things (IoT), a broad class of IoT applications, and how it is potentially relevant in the near future, is presented. In Section 1.2, it is shown how the technology of Narrowband Internet-of-Things (NB-IoT) relates to massive IoT. Section 1.3 provides a brief introduction to the concept of smart grids and the role of massive IoT technologies in the same. Section 1.4 discusses the key motivations for the chosen research subject. Section 1.5 provides an overview of the related work, highlighting the relevant research gaps. Section 1.6 outlines the key research objectives for this study. A high-level overview of the approach used to achieve the stated research objectives is provided in Section 1.7. Finally, the outline of the following chapters is provided in Section 1.8.

1.1 Growing importance of massive IoT

The Internet of Things (IoT) market is expected to cover several industry segments and a wide-range of applications with different Quality-of-Service (QoS) requirements. Figure 1-1 [1] is a depiction of such an IoT application landscape across different industry sectors. The figure shows a partitioning of applications into two broad segments: *massive IoT* and *critical IoT*. This is done to highlight their respective distinguishable application requirements. *Massive IoT*, on one hand, comprises applications involving the exchange of small and infrequent data volumes, and importantly, several stand-alone devices. Hence, for economic viability, such devices must be of low cost and in some cases, support long battery life, e.g. five to ten years. Typically, the latency requirement for such applications is either not specified or is not highly stringent, e.g. in the order of seconds or even up to hours such as for smart metering. Consequently, the network technology providing connectivity to such devices must be aligned with these requirements and support high capacity. *Critical IoT*, on the other hand, consists of time-critical applications, with latency requirements in the order of milliseconds, such as remote surgery and vehicle-to-vehicle (V2V) communications. These applications naturally demand high reliability, availability and low latency.

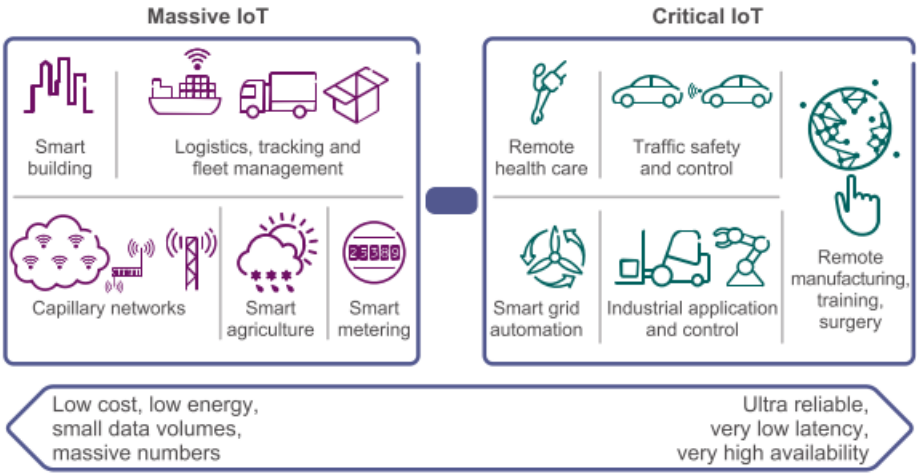

Figure 1-1. Segmentation of the IoT market into massive and critical IoT [1].

Figure 1-2 [1] shows the growth in the overall connected devices, both IoT and non IoT-based, on the internet since 2014 and the predicted numbers from 2017 until 2021. We can see that, by 2021, IoT-based devices will account for 15 billion out of the total 28 billion connected devices.

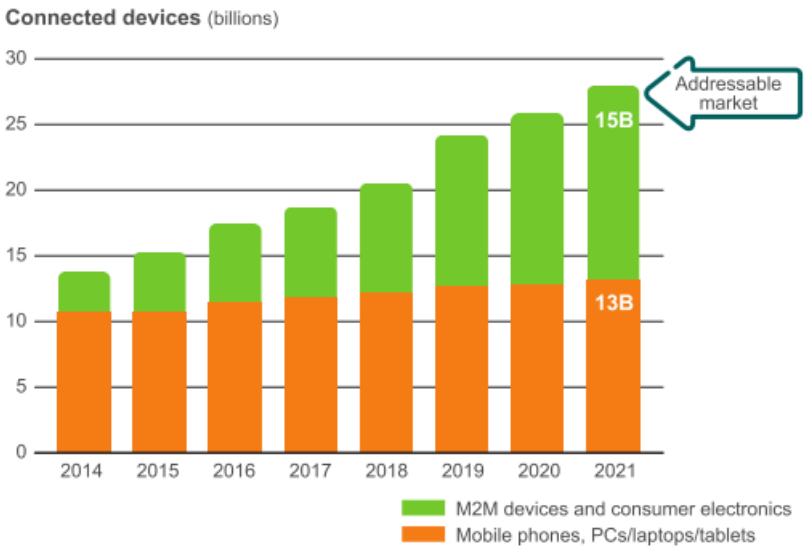

Figure 1-2. Growth in connected devices on the internet [1].

It is expected that in the next four to five years, massive IoT will see more growth in the number of connected devices compared to critical IoT. The reasons for this are two-fold:

1. Low Power Wireless Access (LPWA) technologies such as Narrowband Internet-of-Things (NB-IoT), SigFox [2] and Long Range Wide Area Network (LoRaWAN) [3] are already available to support the needs of massive IoT applications. Hence, this will fuel the introduction of new applications in this segment.
2. Critical IoT is particularly a focus for the 5th Generation (5G) cellular technologies which may see practical adoption only after 2020.

Thus, massive IoT is potentially important in the coming few years which implies that research inputs in this field will be particularly valuable.

1.2 Addressing the needs of massive IoT using NB-IoT

Narrowband Internet-of-Things (NB-IoT) was introduced in July 2016 by the 3rd Generation Partnership Project (3GPP) [4] following a study item [5] in its Release 13 specifications. The objective of this study item was to propose a cellular architecture-based LPWA technology option for low data rate massive IoT applications. Although backward compatibility with legacy LTE technology was not part of the design targets for NB-IoT, some of its design features are indeed derived from LTE and LTE-based enhancements for machine-type communications, e.g. LTE-M introduced in Release 12.

A high-level overview (a detailed discussion is presented in Chapter 3) of the key NB-IoT features is presented below, with emphasis on how these features attempt to achieve the different design targets for massive IoT:

1. **Use of a narrow bandwidth (180 kHz)**
   The bandwidth, each in the uplink and downlink, in NB-IoT is restricted to only 180 kHz (Figure 1-3) which is acceptable for low data rate applications while being able to keep the **complexity** and **cost** of the transceiver low. As shown in the figure, the 180 kHz bandwidth is formed by a set of orthogonal subcarriers as per the Orthogonal Frequency Division Multiplexing (OFDM) principle. Depending on the subcarrier spacing configuration (discussed later), there may be either 12, with a subcarrier spacing of 15 kHz, or 48 subcarriers, with a subcarrier spacing of 3.75 kHz.

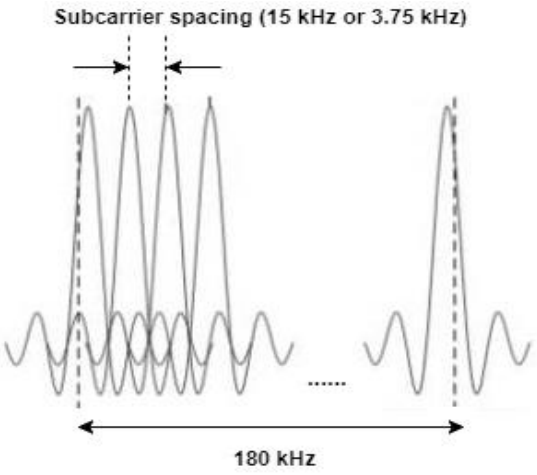

Figure 1-3. NB-IoT bandwidth of 180 kHz consisting of orthogonal subcarriers.

2. **Half-duplex operation**
   An NB-IoT User Equipment (UE) operates in a half-duplex mode which means that the device cannot transmit and receive simultaneously. This further minimises the device **cost** since a duplexer is not needed.

3. **Discontinuous reception and transmission**
   Power Saving Mode (PSM) and extended Discontinuous Reception (eDRx) are two features targeted at minimising **energy consumption** in the end devices, for uplink and downlink-oriented traffic respectively. Essentially, these features allow the UE to transition to a 'sleep' mode for a much longer duration compared to legacy LTE devices.

4. **Repetition in signal transmissions**
   A transport block is the smallest block of data that is transmitted and decoded at the physical layer. In NB-IoT, a transport block can be transmitted multiple times[1] to improve the effective signal strength at the receiver (UE or eNodeB) and hence the signal-to-noise-plus-interference ratio (SINR), which consequently increases the probability of successful decoding. This helps to improve **coverage** in areas of poor received signal strength which may apply to smart meters, since they are expected to be located e.g. in the basements of a building or behind thick walls.

5. **Reduced subcarrier spacing option**
   Transmission in the uplink and downlink are based on the Single Carrier - Frequency Division Multiple Access (SC-FDMA) and Orthogonal Frequency Division Multiple Access (OFDMA) techniques respectively, as is also the case in LTE. However, to improve signal **coverage** in the uplink via Power Spectral Density (PSD) boosting, a subcarrier spacing of 3.75 kHz is available apart from the conventional 15 kHz spacing, as shown in Figure 1-3. Note that in the downlink, only the 15 kHz subcarrier spacing is supported. The choice of the subcarrier spacing to use for a given UE's uplink data transmission can be dynamically made by the operator, e.g. based on the current SINR level of the UE. Consequently, each transmitting UE may have a different subcarrier spacing.

6. **Flexible scheduling of single-tone and multi-tone subcarriers**
   The majority of the applications involving data transmission from a high number of devices are expected to be uplink-oriented, e.g. smart metering and remote waste management. Given the need to support a high number of devices in the uplink, within the narrow 180 kHz bandwidth, the network can flexibly schedule 1, 3, 6 or 12 contiguous subcarriers to individual UEs simultaneously. Further, with the subcarrier spacing of 3.75 kHz, there are 48 subcarriers which can only be allocated

[1] When sending the transport block multiple times, each block may contain the same information bits but not necessarily the same coding bits i.e. those which help in error correction and decoding.

individually, i.e. a single-tone scheduling, to different UEs, which also is aimed to increase **capacity,** in terms of the number of devices that can be simultaneously served.

7. **Flexible deployment options**
   To ensure ease of deployment and coexistence with existing cellular networks (particularly LTE), three modes of deployment are available: in-band, guard-band and standalone. This is depicted in Figure 1-4 [6]. The in-band and guard-band modes allow the NB-IoT carrier (180 kHz bandwidth) to be deployed within an LTE carrier's system bandwidth and guard band, respectively. The stand-alone deployment mode is targeted towards 2G networks where it can be deployed as a 180 kHz carrier within the 200 kHz spectrum previously used for a GSM carrier, with a 20 kHz (i.e. 2x10) guard band.

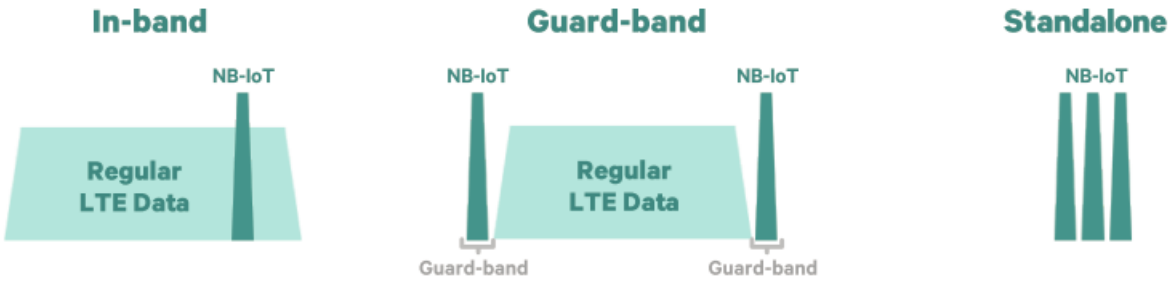

Figure 1-4. The different carrier deployment options in NB-IoT [6].

1.3 Role of massive IoT technologies in smart grids

In this section, the reader is briefly introduced to the concept of smart grids (a detailed discussion is done in Chapter 2) and the role of massive IoT technologies in the same. Based on this, the relevant challenges in smart grids, that form the basis of research in this study, are discussed in the next section.

In broad terms, smart grids are described as energy or electricity networks that can automatically monitor energy flows and adjust to changes in the energy supply or demand accordingly [7]. This is achieved through a two-way communication network that connects certain monitoring and control devices placed across different segments of the power grid i.e. generation, transmission, distribution and customers. At the application layer, one or more applications may utilise this communication network of the smart grid to provide benefits to the utilities, such as high availability, fast detection of and recovery from faults, low peak-demand and high energy efficiency. A conceptual representation of a smart grid in terms of the power grid and the related communication network connecting various monitoring and control devices is shown in Figure 1-5 [8].

In the **power grid**, electricity is first *generated* in bulk quantities through non-renewable and renewable sources and stepped up at the transmission substations to a high voltage (60 – 700 kV). It is then transported through *transmission* lines over long distances. The high voltage power is stepped down to low voltages (200 – 400 V) at distribution substations and delivered via the *distribution* grid to the *customer*s. Note that, in the power grid, certain monitoring and control devices are interfaced, e.g. the smart meter at the customer-end and feeder devices interfaced with the distribution feeder poles/transformers such as switches and reclosers.

The smart grid **communication network** architecture is defined in the IEEE 2030-2011 standard [8] in terms of a correspondence with one or more segments of the power grid described above. There are three important and relevant communication network segments defined – the Home Area Network (HAN), the Neighbourhood Area Network (NAN) and the Wide Area Network (WAN). The *HAN* connects the smart meter to various *customer* appliances, which enables the energy consumption information of these appliances to be communicated to the utilities or third-party service providers, e.g. for energy management or billing services. The *NAN* connects several smart meters and the feeder network with a backhaul network. A feeder network is a local area network which interconnects the feeder devices placed on the *distribution* segment of the power grid. The *WAN* provides connectivity with the utility operations, to the backhaul network and to devices present in the *transmission* substation (transmission substation network), e.g. Phasor Measurement Units or PMUs, and power generation premises (generation network).

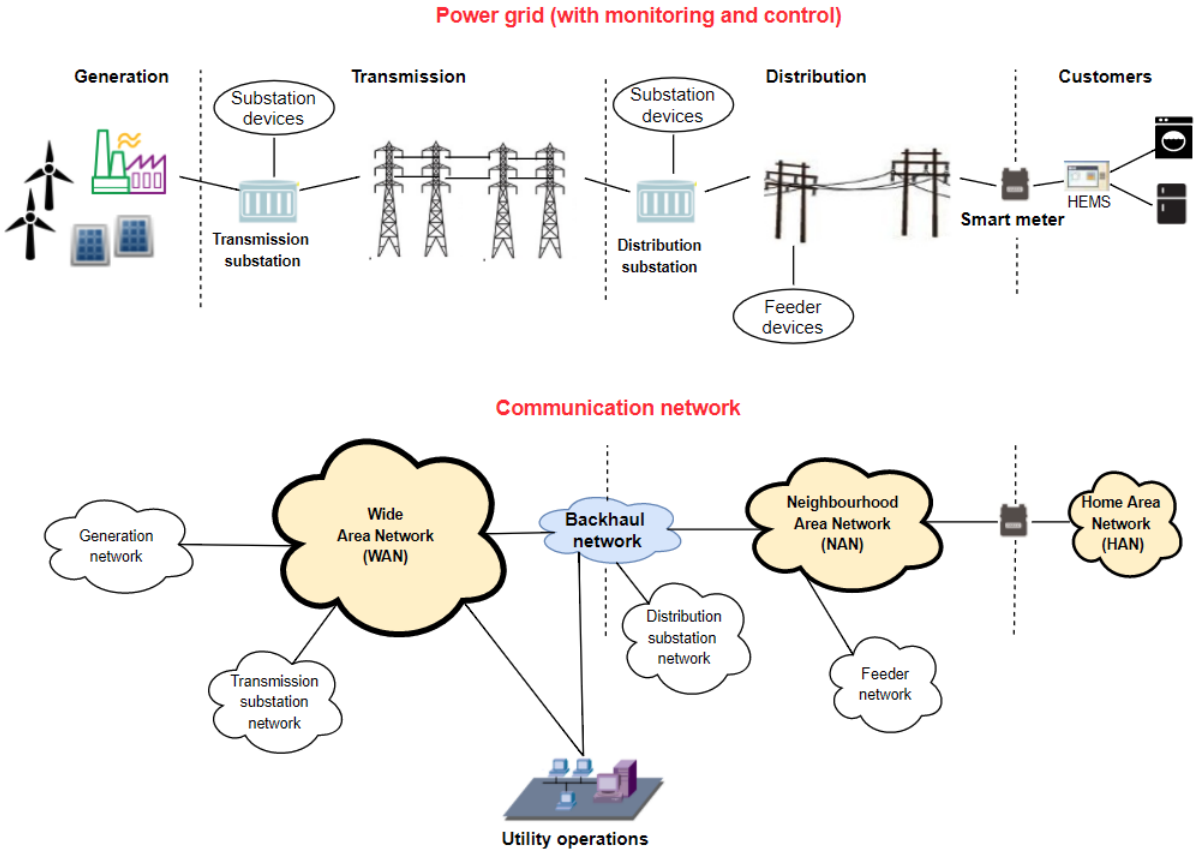

Figure 1-5. A conceptual representation of a smart grid in terms of the power grid and the related communication network connecting various monitoring and control devices [8].

With the help of the WAN, the NAN and the HAN, a two-way communications infrastructure is created, enabling interaction between the utility operations and the monitoring and control devices placed at the different power grid segments. The smart meter is one such monitoring and control device for the customers. The integrated system comprising the smart meters, relevant communications network (e.g. the NAN, the backhaul network) and the monitoring/control applications at the utility operations is referred to as the Advanced Metering Infrastructure (AMI). Massive IoT technologies are generally considered as relevant candidate solutions for the NAN that is part of the AMI. This is because of the need to connect a large number of smart meters with minimal cost of investment and due to the relatively less stringent communication requirements of smart grid use cases that utilize the AMI, as shown later in Chapter 2. Consequently, the focus of this study will be on relevant AMI-related use cases.

1.4 Research motivations

In this section, the key research motivations are discussed, including the relevant challenges in smart grids that we aim to address in this study.

1. **Challenges for massive IoT in the implementation of Outage Restoration and Management (ORM) within the AMI**

   As stated in the previous section, smart meters form an integral part of the AMI. Outage Restoration and Management (ORM) is one of the several important use cases (described later in Chapter 2) for utilities, that is expected to utilise the AMI. It involves smart meters sending a notification, through its communications module i.e. a UE, to an ORM system immediately upon detection of loss or restoration of power [9]. Such notifications are a cost-effective alternative to dependence on customer calls to detect power outage affected areas. Furthermore, if a sufficient number of notifications from smart meters are received soon enough, they can help ORM systems determine the

power outage location (power outage localisation) more quickly and accurately, and perform fast restoration.

ORM, unlike the other use cases in AMI, is particularly challenging from a communications network point-of-view. Depending on the extent of the power outage, i.e. the number of affected households/smart meters, the network may have to deal with 'near' simultaneous access by several UEs, leading to a suddenly high load on and potential congestion of the physical communication resources i.e. the wireless channels. Consequently, several of the notification messages may fail to be delivered successfully or within a certain end-to-end delay target.

In the context of ORM, success rate is the percentage of the total generated notifications that have been successfully delivered to the ORM system. The success rate of the notifications that arrive within a certain end-to-end delay target is defined in terms of a metric called *reliability* [9]. The end-to-end delay target for ORM is specified as 20 seconds in the literature [9]. On the other hand, there is variation in the general minimum reliability requirement specified in the literature. For instance, in [9], a minimum reliability of 30% for large power outages is specified whereas, in [10] and [11], reliability requirements of 98% and 99 % are stated respectively. This is likely due to the fact that the minimum success rate of notifications required to guarantee sufficient accuracy of power outage localisation, may depend on the extent of the power outage, the location of the affected smart meters and the distribution grid topology, as can be concluded from [12] and [13]. These studies have stated that the higher the reliability, the higher is also the accuracy of the power outage localisation. Thus, for a given number of power outage-affected smart meters, the key challenge for the network is to maximise the success rate of the notifications while ensuring that the delay target is met, in order to maximise the reliability. Note that the delay target of 20 seconds could be challenging for massive IoT technologies. For instance, although it has been shown that NB-IoT can meet an uplink delay target of 10 seconds [14], the traffic model of use cases like ORM were not specifically considered in the evaluation. Such use cases have been shown in existing studies to have a profound impact on the end-end delay even in technologies like LTE. For example, it has been shown in [15] that, for a use case similar to ORM, the LTE uplink end-end delay, originally designed to be within 100 ms in the worst-case, can lie between 70 milliseconds to 1 second on average depending on the load, in order to achieve at least a 30% success rate.

2. **Potential relevance of the research**

In this thesis, NB-IoT is chosen as a candidate massive IoT technology solution for evaluation with respect to the ORM use case in smart grids. A research in this direction would be relevant because of the expectation that NB-IoT may be a more relevant technology for both utilities and network operators in the context of deployment in smart grids. Furthermore, the results of this thesis would aid the current efforts towards enhancements to massive IoT technologies in general. These aspects are further explained below:

i. In comparison to unlicensed technologies in the LPWA space, such as SigFox [2] and LoRa [3], 3GPP based technologies are specifically suited for wide-area deployments for large businesses/industries such as utilities. In such cases, the benefits of reaching a large customer base and flexibility to offer a wide-variety of applications with different QoS and security requirements, may outweigh the costs of using a licensed technology. Furthermore, network operators, equipment vendors, chip manufacturers and researchers around the world have been continuously developing 3GPP standards and related hardware chipsets over the past two decades and this broad industry support is expected to continue. Hence, 3GPP technologies are in some ways 'future-proof'.

ii. As stated in Section 1.2, NB-IoT offers flexible deployment options such as 'in-band' and 'stand-alone', which helps mobile network operators (MNOs) reuse their existing networks to offer extra services without additional licensing costs.

iii. In the context of smart grids, MNOs can offer interesting business models [16] to Distribution Service Operators (DSOs) wherein the end-to-end communication infrastructure is setup/managed by the MNO and linked to the company data centres. This saves the utilities the cost of operating and maintaining their private networks. Furthermore, service level agreements, security and billing could be added to the business model for offering ''NB-IoT As a Service" [17].

iv. The three use case categories or usage scenarios, specified by the International Telecommunications Union-Radio (ITU-R) for the 5G wireless standards (termed as IMT-2020), are shown in Figure 1-6 [18]. These are *enhanced mobile broadband*, *massive machine-type communications* (massive IoT) and *ultra-reliable and low-latency communications* (critical IoT). In the 5G wireless standards currently being developed by 3GPP, NB-IoT is being used as the foundation for further enhancements to meet the requirements of the massive IoT use cases. Hence the results and learnings from this study will be a useful input for researchers and industry stakeholders working on these enhancements.

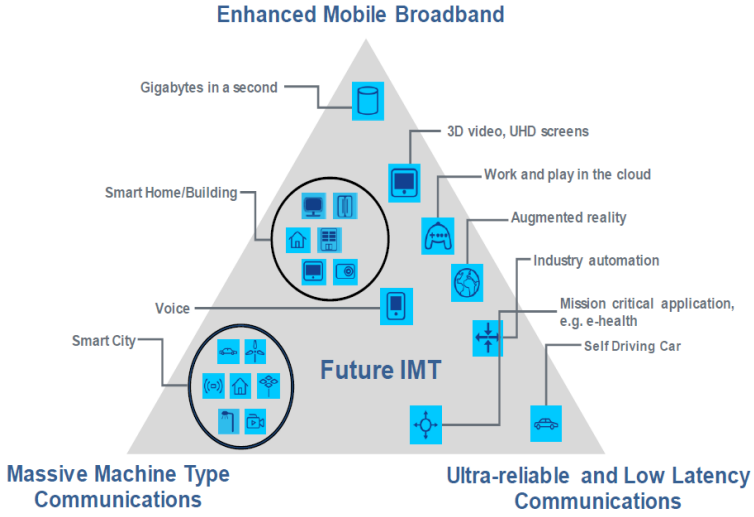

Figure 1-6. ITU-R usage scenarios for 5G [18].

3. **Need to analyse the inherent trade-offs within NB-IoT**

In Section 1.2, the key features of NB-IoT were discussed and linked with the massive IoT design targets. It is interesting to note that there are inherent trade-offs associated with these features. For example, the reduction in the available bandwidth implies that the transmission time may need to be increased to send the same amount of data. Thus, the scheduling of single-tone subcarriers, or the use of a 3.75 kHz subcarrier spacing, increases the number of UEs that can be served simultaneously, i.e. the scheduling capacity, but it may also increase the overall transmission time which may bring a trade-off between the scheduling capacity and latency (or energy consumption). Similarly, an increase in the transport block repetitions improves the coverage for a UE. However, the UE would then occupy resources for that much longer and experience a higher latency in its transmission, which expresses a trade-off between coverage and latency. The main challenge, when handling massive network access such as that in a power outage scenario is to balance these mutual trade-offs to maximise the overall reliability performance. The design of the scheduler would play a key role in addressing this challenge. Note that the time-frequency resources allocated to a UE by an NB-IoT scheduler must, primarily, include the start time of the transmission, the number of allocated subcarriers, the assigned subcarrier spacing and the number of transport block repetitions.

1.5 Related work

NB-IoT is a relatively new technology standard. Thus, the extent of available contributions on this topic is rather limited. Much of the existing work on NB-IoT ([19]- [24]) has focussed on analysis and enhancements to certain technology elements such as paging mechanisms and data transmission protocols, and not addressed the performance impacts due to use cases such as ORM, involving massive network access.

In [25] and [26], coverage and latency results for NB-IoT are presented. However, the traffic model considered is a generic one and not specifically targeted at smart grid type use cases. A simplistic fixed scheduling scheme, with twelve or single tone allocation with 15 kHz subcarrier spacing, was used and the impact due to different deployment modes, multi-tone subcarrier allocation or lower subcarrier spacing was not studied. Moreover, the effect of different traffic loads was also not considered.

The authors of [27] present a capacity analysis for NB-IoT in urban and suburban environments, considering smart metering applications. However, the results are based on analytical calculations with optimistic data rate assumptions i.e. without consideration of system-level overheads and limitations in the choice of the Modulation and Coding Schemes (MCSs).

In [28], a new concept of control channel load balancing for NB-IoT is introduced which is aimed at a dynamic allocation of control channel resources, based on sudden changes in traffic load, amongst UEs in different coverage levels in NB-IoT (the concept of coverage levels is explained in Chapter 3). Although the concept is relevant to the current research problem, the proposed methodology is based on pre-standard specifications and thus, the methodology would need to be adapted as per the current standard.

Certain LTE-based schedulers targeted for machine-type communications have been proposed in the literature, with the aim to maximise packet success rate and minimise end-to-end delay. Similar principles may be applicable, though with certain adaptations in the design of the scheduler in this study. In ([29]- [32]) certain Earliest Due Date First (EDDF) based and channel-dependent LTE schedulers have been shown to be effective in maximising the packet success rate and minimising the 95th percentile delay. The key idea is to prioritise UEs based on a metric which is dependent on the waiting time of the Head-of-Line (HOL) packet at the UE's buffer, the packet delay budget and the wireless channel quality experienced by the UE. The challenge when considering uplink oriented traffic is that the waiting time of each packet at the UE buffer is not accurately known by the scheduler since the Buffer Status Report (BSR) in LTE and NB-IoT only conveys the amount of data in the buffer and not the time spent in the buffer by individual packets. Hence, for practical adoption, certain approximations need to be made for estimating the waiting time.

In [33], an EDDF based scheduler for uplink traffic is proposed, to address the challenge of the waiting time calculation discussed above. However, it involves additional signalling overhead to convey the waiting time of the HOL packet, which requires a change in the current 3GPP specifications.

1.6 Research objectives

Based on the research motivations and related work discussed in the previous section, the research objectives of this thesis can be stated as follows:

1. Assessment of the suitability of NB-IoT as a communications technology for smart meters to facilitate ORM in various realistic deployment and power outage scenarios.
2. Optimisation of the network configuration of NB-IoT, with a specific focus on the packet scheduler, in order to maximise the reliability performance of ORM.

1.7 Approach

The step-wise approach (high-level) used to achieve the research objectives stated in Section 1.6 is outlined as follows:

1. Identify the key application characteristics and communication requirements of relevant AMI-based use cases.
2. Obtain a detailed understanding of 3GPP's NB-IoT specifications, particularly those addressing the physical (PHY) and Medium Access Control (MAC) layers.
3. Obtain a qualitative understanding, of the NB-IoT deployment/configuration options and the associated trade-offs.
4. Study the state-of-art in system-level simulation of mobile networks and smart grid-related modelling.
5. Develop and implement a system-level simulation model that would enable performance analysis of ORM-based network traffic (modelled based on Step 1) for different NB-IoT deployment and network configuration (modelled based on Steps 2, 3 and 4) scenarios.
6. Design candidate NB-IoT schedulers to be integrated with the simulation model in Step 4 and later, for evaluation.
7. Using the simulator developed in Step 5, carry out a performance evaluation for ORM-based network traffic under different scheduler, network deployment, configuration settings and power outage scenarios.
8. Derive key conclusions and recommendations based on the obtained results in Step 6.
9. Provide recommendations for the further extension of the work, based on the general conclusions/insights and limitations, if any, in the current work.

1.8 Thesis outline

The remainder of this thesis is as follows. An overview of the key concepts associated with smart grids and the application characteristics and communication requirements of AMI-based use cases, including ORM, is provided in Chapter 2. In Chapter 3, the relevant specifications of the NB-IoT technology are described. Chapters 2 and 3 will form a technological basis for the simulation model described in Chapter 4 and the candidate NB-IoT schedulers proposed in Chapter 5. The simulation scenarios, results and discussions are presented in Chapter 6. Finally, the key conclusions and recommendations are provided in Chapter 7, along with a suggested direction for future research.

Chapter 2. Overview of smart grid communications and requirements

2.1 Introduction

Since the broad theme of this study deals with communications in smart grids, it is important to have a clear understanding of the key concepts, motivations, underlying architecture and requirements. Thus, in this chapter, the following questions are addressed:

1. *What is a smart grid and how is it expected to be better than the conventional power grid? (Section 2.2)*
2. *Why are communications important for a smart grid? (Section 2.3)*
3. *What are the general traffic characteristics and communication requirements of the smart grid use cases relevant for massive IoT technologies such as NB-IoT? (Section 2.4 - 2.5)*

2.2 Concept of a smart grid

A smart grid is defined in [34] as an ''intelligent power network that combines various technologies in power, communication, and control, which can monitor and optimise the operations of all functional units from electricity generation to end-customers". More formally, the IEEE 2030-2011 standard [8] defines a smart grid as an automated, widely distributed energy delivery network characterised by a two-way flow of electricity and information, capable of monitoring and responding to changes in everything from power plants to customer preferences to individual appliances. Regardless of the many different 'formal' definitions of a smart grid, the common acceptance is that it will involve a two-way flow of electricity and information to enable automated and distributed energy delivery.

Compared to a conventional power grid, a smart grid is expected to provide the following important benefits to utilities and energy consumers [35]:

1. Improved reliability.
2. Increased physical, operational and cyber security and resilience against attack or natural disasters.
3. Ease of repair and recovery during power outages/faults.
4. Increased information available to consumers regarding their energy usage.
5. Increased energy efficiency and sustainable energy usage.
6. Integration of greater proportion of renewable energy sources.
7. Integration of plug-in electric vehicles and energy storage devices.
8. Reduction in peak demand and resulting increase in revenue of utilities.

2.3 Communications in a smart grid

A key component of a smart grid is the existence of a communication network that covers the different segments of the power grid, as introduced in Chapter 1, which facilitates a two-way exchange of information. The exchange of information typically occurs between applications in the utility operations and certain monitoring and control devices in the different power grid segments. These elements, to name a few, include sensors and switches at the distribution substations, Phasor Measurement Units (PMUs) in the transmission segment, and smart meters at homes/industries. As introduced in Figure 1-5, the smart grid communication network is divided into three hierarchical segments, the HAN, the NAN and the WAN, which also correspond to one or more segments of the power grid. Figure 2-1 includes a similar representation, except with the NAN shown by a generic network of smart meters and so-called 'aggregators', and the backhaul network shown by a cellular network. Note that in the figure, for clarity, only few relevant interconnections/interfaces between the main network segments and local networks, e.g. the feeder network and utility operations, have been shown, while more possible interfaces are defined in the IEEE 2030-2011 standard.

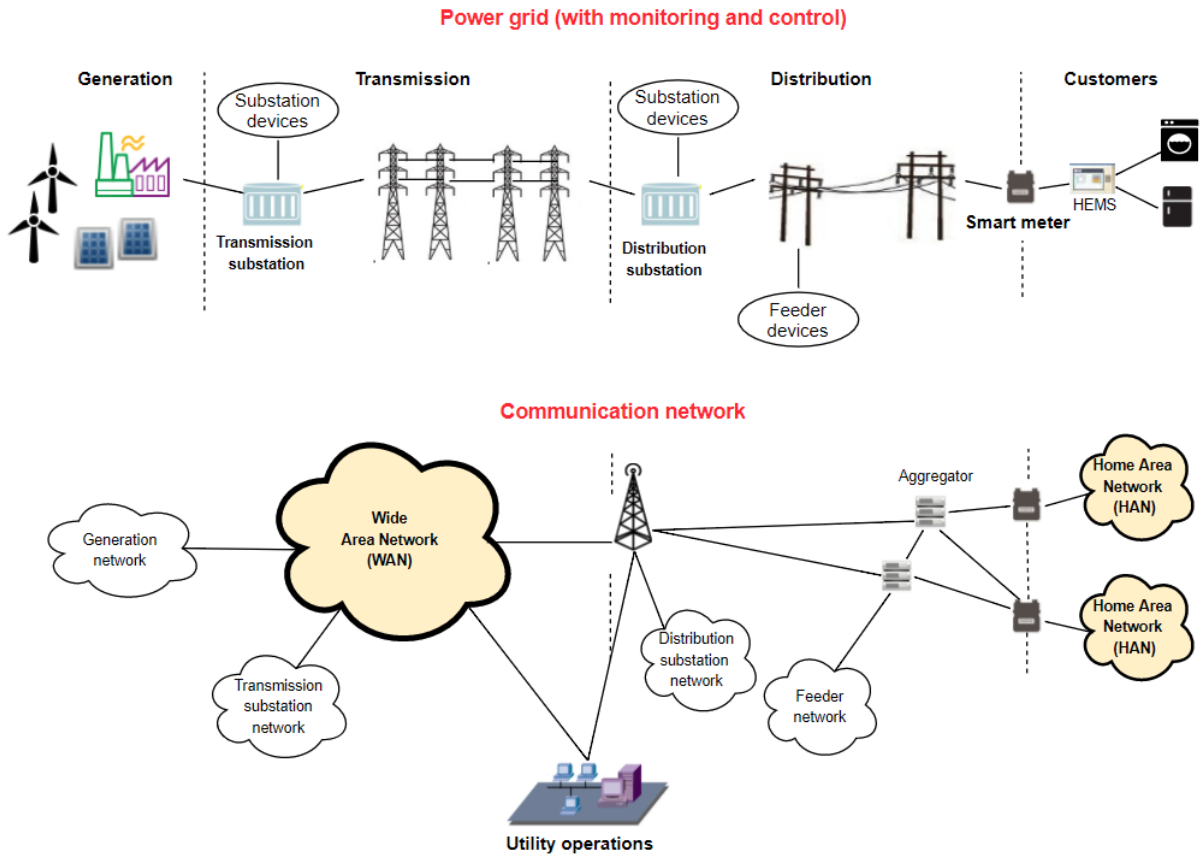

Figure 2-1. Representation of the hierarchical smart grid communication network in relation with the different power grid segments and various monitoring and control devices.

In relation to the figure, the three network segments are explained in further detail as follows:

1. **HAN**

   The HAN is typically associated with the *customer* segment of the power grid. It is a network connecting different customer appliances, Distributed Energy Resources (DERs) such as solar panels, and Home Energy Management Systems (HEMS) to the smart meter. Most importantly, this enables the smart meter to record the energy consumption trend for different appliances and allows utilities or third-party service providers to deliver energy usage data (e.g. on the HEMS) to the customers and load control commands to the appliances, e.g. postponing the wash cycle of a washer, via the HEMS. The coverage area of a HAN can be upto 100 m [10].

2. **NAN**

   The NAN is typically associated with the *distribution* segment of the power grid. It connects several smart meters and feeder networks to the backhaul network. A feeder network is a local area network which connects distribution feeder devices such as switches, reclosers and feeder sensors [36]. The coverage area of a NAN can be between 100 m to 10 km [10] .

   Note that the aggregator unit shown in the figure, which aggregates data from a set of smart meters, is a generic representation. In practice, deployments follow both 'aggregator' and 'non-aggregator' approaches, depending on the technology supported by the smart meter communication module. Typically, aggregators are required when using short-range or mesh-based communication technologies like PLC and IEEE 802.15.4g. They may not be required, thus avoiding the corresponding investment costs, if the communication modules support long-range cellular technologies e.g. GSM and NB-IoT. Hence, only a non-aggregator deployment is modelled in this study. Consequently, in this case, the cellular network can be considered as part of the NAN.

3. **WAN**

   The WAN is typically associated with the *transmission* and *generation* segments of the power grid. It primarily connects the local generation and transmission substation networks with the utility operations. Additionally, it may also link the backhaul network with the utility operations. Within the transmission substation networks, there are devices such as PMUs and Intelligent Electronic Devices (IEDs). PMUs, for example, collect power measurement samples (current and voltage phasors) at high frequencies (around 30-60 Hz), typically from the transmission lines or at transmission substations, for purposes of fault localisation/protection and state estimation of the grid. The coverage area of a WAN can be between 10 km to 100 km [10].

2.4 General communication requirements of smart grid use cases

Several smart grid use cases exist, that exploit the communication and control layer described in the previous section, and collectively provide utilities with the desired benefits stated in Section 2.3. These use cases can be broadly classified based on the power grid segment they are primarily associated with. Thus, the use cases are classified as *transmission and generation segment-based*, *distribution-segment based* and *customer segment-based*. Table 2-1 [10] lists some example use cases for these categories, few example power grid components involved and the typical end-end latency requirements. Note that the end-to-end latency requirement is for the message transfer between the relevant power grid components and the utility/third-party operations.

Table 2-1. Example smart grid use cases associated with the different power grid segments [10].

Use case category	Example power grid components involved	Example use cases	End-to-end latency requirement (order of magnitude)
Customer segment-based	Customer appliances, HEMS, DERs	Home automation, building automation	seconds
Distribution segment-based	Feeder devices, smart meters, distribution substations	Meter reading, Real-Time Pricing (RTP), Outage Restoration and Management (ORM)	seconds to hours
Transmission and generation segment-based	PMUs, transmission substations, generators	State estimation, voltage stability control	milliseconds to minutes

From Table 2-1, we can see that the *transmission and generation segment-based* use cases have the most stringent communication requirements, with most use cases requiring an end-to-end latency in the order of milliseconds, e.g. 100 ms in the case of state estimation. Consequently, communication solutions supporting high data rates, such as LTE, WiMAX and Fiber are typically deployed, as a WAN, to address the needs of such use cases [10][37]. *Customer segment-based* use cases such as home automation have relatively relaxed communication requirements. Since the coverage area is restricted to the customer premises, short-range communications solutions such as Ethernet, ZigBee and Bluetooth are typically deployed [10][37]. Use cases such as meter reading, involving communication between the utility and the smart meter, over the AMI, are classified as *distribution segment-based*. Due to the need to provide low-cost connectivity to several smart meters and the relatively less stringent communication requirements, massive IoT technologies are considered more relevant than other technologies for such use cases.

In the following section, we will look at the relevant distribution segment-based smart grid use cases, their traffic characteristics and communication requirements.

2.5 Distribution segment-based use cases: traffic characteristics and communication requirements

In this section, the traffic characteristics and communication requirements of distribution segment-based use cases relevant to this thesis are presented.

A standard specification for the traffic characteristics and requirements for distribution segment-based use cases is not available. Instead, there is a variation in the traffic models that have been presented or assumed in the available work in the literature. However, the most comprehensive and widely cited source of traffic patterns and network requirement specifications for distribution segment-based use cases is that provided by the OpenSG task force [9], which has been used in this thesis. The OpenSG task force is a consortium of 190 companies which includes utility companies, network operators and consultancy organisations. As per [38], the use cases considered by this consortium are consistent with those studied by standardisation organisations such as the European Telecommunications Standards Institute (ETSI) and Universal Smart Energy Framework (USEF). The most important distribution segment-based use cases included in the OpenSG specifications can be classified within the below broad categories [9][10]:

1. Advanced Metering Infrastructure (AMI)
2. Demand Response (DR)
3. Distribution Automation (DA)
4. Distribution Customer Storage (DCS)
5. Electric Transportation (ET)

AMI is an integrated two-way communications and management system that allows the utilities to collect, measure and analyse energy usage data from smart meters, on-demand/event-triggered or on a pre-defined schedule, for power outage management, grid management and billing purposes [35][39]. A typical AMI use case is meter reading.

DR is the reduction in energy usage by customers or at the customer-end in response to higher energy price or increase in load [35]. DR is quite closely related to AMI since AMI is typically the underlying infrastructure needed to enable DR and there are some applications classified in either category by researchers and the industry. A typical DR use case is a load-control mechanism implemented by the utility or a third-party service provider. This involves the utility remotely controlling the energy usage of an appliance at home, e.g. by controlling the air-conditioner temperature setting, during times of peak energy demand or dynamic changes in the energy pricing. Here, the smart meter acts as a gateway for the utility to control the appliance.

DA use cases allow the utility to monitor and control the status of feeder devices (see Figure 2-1) such as switches, reclosers and feeder sensors [9]. A typical DA use case such as Distribution System Monitoring and Maintenance (DSMM) involves periodic polling of the data from sensors integrated with the distribution grid devices, for real-time monitoring of their status and early detection of faults.

DCS use cases allow efficient integration of electric grid storage devices installed on the distribution feeder circuits for purposes of reduction of peak demand, power quality maintenance and power interruption protection for the feeder circuit itself and/or a set of customers served by the feeder circuit [9]. A typical DCS use case involves the utility sending discharge commands to the grid storage device, at periods of high load followed by a charge command during low-load periods.

ET use cases allow electric vehicles, such as those based on battery, hybrid and fuel cells, to act as mobile DERs [9]. Thus, it may involve electricity flow from vehicles to the power grid (vehicle-to-grid) or vice-versa (grid-to-vehicle). A typical ET use case may include a two-way communication between the utility and the vehicle in which the vehicle sends the charge-status of its battery in response to a request from the utility.

In this study, we are mainly interested in the smart meter-related use cases since they are most relevant for massive IoT technologies. Hence, the traffic characteristics and requirements of the first two categories, shown earlier, are discussed in the further subsections. For the remaining categories, interested readers may

refer to [9]. In the following two subsections, the underlying use cases in AMI and DR, their traffic patterns (direction of transfer, occurrence frequency and payload sizes) and communication requirements are described.

2.5.1 Advanced Metering Infrastructure (AMI)

Table 2-2 summarises the general traffic characteristics of AMI based use cases, based on the data available from [9].

Table 2-2. Traffic characteristics and communication requirements for AMI.

Use case	Traffic flow	Event frequency	Payload (bytes)	End-end latency requirement	Reliability requirement
Meter reading – On Demand	Utility → Meter (read command) Meter → Utility (response)	25 per 1000 meters per day	25 (read) 100 (response data)	< 15 s	> 98 %
Meter reading- Scheduled	Meter → Utility	1 per meter per 4 or 6 hours (residential meters) 1 per meter per 1 or 2 hours (industrial meters)	1600 – 2400 200 – 1600	< 4 hours < 2 hours	> 98 %
Meter system events	Meter → Utility	4 per 1000 meters per month	278	< 4 min	> 98 %
Outage Restoration and Management (ORM)	Meter → Utility	1 per meter per event (power lost or restored)	25	< 20 s	> 30 % for large outages[2] (No target specified for small outages)
Real-Time Pricing (RTP)	Utility → Meter Meter → Utility (ACK)	60 per 1000 meters per day	100 25 (ACK)	< 5 s	> 99 %
Time-of-Use Pricing (TOU)	Utility → Meter Meter → Utility (ACK)	60 per 1000 meters per day	100 25 (ACK)	< 5 s	> 99 %
Service switch operation	Utility → Meter Meter → Utility (ACK)	1-50 per 1000 meters per day (8AM – 8PM)	25 25 (ACK)	< 1 min < 2 min	> 98 %
Firmware updates	Utility → Meter Meter → Utility (ACK)	2 per meter per year	400K – 2000K 50 (ACK)	< 4 hours < 5 s (ACK)	> 98 %

[2] The definition of what constitutes a 'large' and a 'small' power outage, in the terms of the range of number of affected smart meters/substations, is not provided.

Overall, there are two general categories of traffic: meter-initiated traffic and utility-initiated traffic. It can be observed that the event frequencies are, in general, very low with the highest frequency of one per hour for scheduled meter readings. For ORM, with every occurrence of power outage or restoration, every affected smart meter sends an uplink notification message to the utility. Hence, it may involve 'near' simultaneous network access from several devices, depending on the extent of the power outage. Moreover, the end-to-end delay requirement for ORM is relatively stringent (20 seconds). Although the delay requirement for certain other use cases is even more stringent, e.g. five seconds for TOU, the event frequency indicates that it is less likely to impose significant instantaneous load arising from multiple devices sending data around the same time such as for ORM. Hence, in the AMI category, **scheduled meter reading and ORM** can be considered the most demanding for an NB-IoT network, in terms of average/instantaneous network load and delay requirements.

2.5.2 Demand Response (DR)

The traffic characteristics and communication requirements for DR are shown in Table 2-3. The main use case in DR is the Direct Load Control (DLC) in which the utility sends a load control command, for a customer appliance, to the smart meter. The smart meter then forwards this command to the appliance over the HAN (see Figure 2-1). The smart meter sends an acknowledgement response once the command is successfully executed. Similar to most use cases in AMI, the event frequency and network load is low.

Table 2-3. Traffic characteristics and communication requirements for DR.

Use case	Traffic flow	Event frequency	Payload (bytes)	End-end latency requirement	Reliability requirement
Direct Load Control (DLC)	Utility → Meter Meter → Utility (ACK)	60 per 1000 meters per day	100 25	< 5 s	> 99 %

2.6 Summary

In this chapter, we have looked at the basic concept of a smart grid, its architecture and the benefits offered over a conventional power grid. A key idea in the vision of a smart grid is to have two-way communications infrastructure between various power grid segments and the utility, enabling the balance of energy demand/supply, early detection/recovery of faults and various other use cases. Hence, the performance of the communication network is a vital component in the operation of a smart grid. We have seen that the communication requirements vary between different categories of smart grid use cases, with the transmission and generation segment-based use cases being the most stringent. Further, it was seen that the distribution segment-based use cases which involve utility-smart meter communications are the most relevant for massive IoT technologies. Finally, the detailed traffic characteristics and communication requirements of distribution segment-based use cases were explored, based on the specifications of the Open SmartGrid (OpenSG) task force. The focus was on smart meter-related use cases in the categories of AMI and DR. Amongst the presented use cases, scheduled meter reading and ORM could be considered, relatively, the most demanding for the network, in terms of average/instantaneous network load and delay requirements. The insights from this chapter will be used in the modelling phase in Chapter 4, particularly with regards to the network layout and traffic models for simulation.

Chapter 3. Overview of the NB-IoT technology

3.1 Introduction

This chapter intends to provide a technological introduction to NB-IoT, focusing on the key motivations and trade-offs behind specific design choices and present only those technical aspects that are essential, to derive the models discussed in Chapter 4, and for the scheduler design in Chapter 5. Thus, appropriate references, mainly to the Release 13 3GPP specifications, are provided, where necessary, for interested readers to obtain more detailed information.

The outline of this chapter is as follows. Section 3.2 describes the key features of the air interface and Section 3.3 provides an overview of the physical layer resource structure. Next, the implementation of scheduling at the physical layer is discussed in Section 3.4, which provides the technological basis for the design of the scheduler in Chapter 5. Since the traffic models in this study involve small payloads, in Section 3.5 we look at the key steps involved in an efficient, i.e. with a low signalling overhead, small data transmission in NB-IoT. The chapter concludes with a summary in Section 3.6.

3.2 Air interface: key features

The air interface mainly represents the set of physical layer specifications required to communicate, via a user equipment or UE, with the access network of NB-IoT. First, an overview of the design of this access network is presented in Section 3.2.1. Then the important air interface specifications for an NB-IoT UE are discussed in Section 3.2.2. Section 3.2.3 describes the so-called 'multi-carrier operation' introduced in NB-IoT. Finally, in Section 3.2.4, the distinct carrier deployment modes available in NB-IoT are discussed briefly.

3.2.1 Access network design

The overall access network design is shown in Figure 3-1 [40].

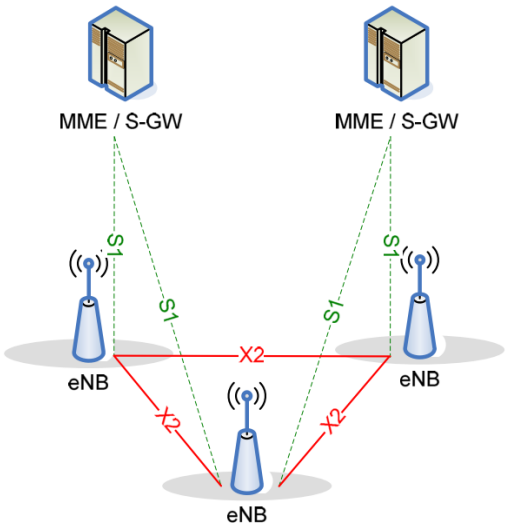

Figure 3-1. NB-IoT access network design [40].

In principle, the access network is represented only by the network of eNodeBs/eNBs (base stations) linked by the logical interface X2. For completeness, the access network interface with the core network is also shown, via the S1 interface linking the Mobility Management Entity (MME[3]) and Serving Gateway (S-GW[4])

[3] MME is responsible for functions such as paging and authentication of the UEs.

[4] S-GW is responsible for functions such as routing of data packets.

nodes, both of which are part of the NB-IoT core network. Note that this access network design is used in LTE as well.

3.2.2 UE specifications for the air interface

The UE category applicable to NB-IoT is named as 'Cat-NB1.' Table 3-1 summarises the key specifications [42] for a Cat-NB1 UE, which were also briefly introduced in Section 1.2. We can easily see why the term 'narrowband' in NB-IoT has been used. It is because the UE supports only a narrow carrier bandwidth of 180 kHz, which ultimately leads to a reduction in the device complexity and cost compared with that of wideband devices. Note that this bandwidth limit applies to both the downlink and the uplink.

The downlink transmission is based on OFDMA with a 15 kHz subcarrier spacing as is also the case in LTE. The uplink design is slightly different to that of LTE. Although SC-FDMA is still the multiple-access scheme used, to support low-complexity UEs and a high number of simultaneous access, a new smaller subcarrier spacing of 3.75 kHz is introduced in the uplink in addition to the normal 15 kHz subcarrier bandwidth. The detailed time-frequency resource structure is discussed later in Section 3.3.

Table 3-1. A summary of the Cat-NB1 UE specifications [42].

Parameter	Value
Bandwidth	180 kHz
Subcarrier spacing	Downlink (OFDMA): 15 kHz Uplink(SC-FDMA): 15 kHz or 3.75 kHz
Operating bands	800 – 850 MHz, 900 MHz and 1800 MHz
Max output power	23 dBm (Class 4), 20 dBm (Class 5)
Duplex operation type	Half-Duplex Frequency Division Duplexing (Type B)
Peak throughput	Downlink: 227 kbps Uplink: 250 kbps

Table 3-2 [40] summarises the transmission configuration options for the 15 kHz and 3.75 kHz subcarrier spacing for both the downlink and the uplink. In the downlink, transmission occurs with twelve subcarriers (multi-tone) in the frequency domain. Note that a 'tone' here refers to a subcarrier. In the uplink, there is some flexibility in the transmission bandwidth, depending on the subcarrier spacing. With the 3.75 kHz subcarrier spacing, transmission with only one subcarrier (single-tone) is supported whereas, for the 15 kHz subcarrier spacing, single and multi-tone transmission are both possible. In the multi-tone case, transmission with three, six or twelve subcarriers is supported.

One of the motivations for transmission with less than twelve subcarriers is to serve UEs in poor coverage which cannot profit from having a high bandwidth. This may be an important requirement for devices such as smart meters which are generally located in poor signal coverage conditions such as behind walls or in basements. Decreasing the bandwidth increases the SINR through PSD boosting, enhancing the likelihood that UEs can decode the received signals correctly. However, the trade-off could be a longer transmission time for the same amount of data. But this may be acceptable if the target applications are relatively delay-tolerant.

From Table 3-2, we can further observe that modulation types higher than QPSK are not supported. Given that this helps to minimise chipset complexity (hence cost) and data rates of targeted applications are expected to be low, a loss in spectral efficiency due to the absence of higher order modulation schemes in good SINR conditions may be an acceptable trade-off. Another observation is that, for a single-tone transmission, $\pi/2$-BPSK and $\pi/4$-QPSK are supported instead of plain BPSK and QPSK. This is done in order to minimise the peak-to-average power ratio, which increases the efficiency of the power amplifier [26]. The trade-off is a coding loss i.e. a higher SINR is needed to attain the same Block Error Rate (BLER).

Table 3-2. Supported transmission configuration for different subcarrier spacing options [40].

Direction	Subcarrier spacing (kHz)	Type of transmission	Transmission bandwidth (kHz)	Multiple-access scheme	Modulation scheme
Downlink	15	Multi-tone (12 subcarriers)	180	OFDMA	QPSK
Uplink	3.75	Single-tone	3.75	SC-FDMA	$\pi/2$-BPSK, $\pi/4$-QPSK
	15	Single -tone	15	SC-FDMA	$\pi/2$-BPSK, $\pi/4$-QPSK
		Multi-tone (3, 6 or 12 subcarriers)	45, 90 or 180		QPSK

The operating bands (Table 3-1) for NB-IoT are a subset of the list of bands for LTE [42] in the 800, 900 and 1800 MHz range. Most of the available bands are in the lower frequency range (800 – 900 MHz), to improve coverage.

Unlike the previous UE categories in the 3GPP releases, Cat-NB1 supports only a Half-Duplex Frequency Division Duplexing (HD – FDD) operation [43]. The 'FDD' part means that a pair of carriers with a certain frequency separation are deployed for downlink and uplink transmission. The 'HD' part means that the UE cannot receive and transmit simultaneously on the carrier pair. A half-duplex operation helps to minimise device cost and complexity since it enables the reuse of Radio Frequency (RF) hardware for transmission and reception, e.g. the use of a single oscillator, and eliminates the need for a duplexer. Since simultaneous uplink transmission and downlink reception is not possible, a guard period is required so that the transceiver can switch between reception and transmission. As mentioned in Table 3-1**,** NB-IoT supports a Type B HD-FDD operation. In this type of operation, the guard period lasts for the duration of a 'subframe' in the downlink [44]. The concept of a subframe is introduced later in Section 3.3.

Certain functions typically supported by LTE UEs (Release 8 and beyond) are not supported for an NB-IoT UE in the current release, with the aim of minimising the UE complexity. Some of these key functions are inter-RAT and intra-RAT handover, Channel Quality Indicator (CQI) reporting, Multimedia Broadcast Multicast Services (MBMS) and Carrier Aggregation (CA).

3.2.3 Multi-carrier operation in NB-IoT

In a generic deployment scenario, the pair of carriers deployed in the downlink and the uplink are used for both signalling and data transmission purposes. Signalling functions in the downlink include the initial cell search, synchronisation, broadcast information and paging whereas in the uplink, it includes primarily the preamble transmission for random access (see Section 3.5). These signalling functions may leave a restricted amount of resources (time and frequency) for data transmission. To address this issue, *multi-carrier operation* is supported [40]. In this type of operation, one or more pairs of secondary carriers can be configured purely for data transmission, including the associated scheduling messages. In such a scenario, the primary carrier, in the downlink or uplink, used for signalling is referred to as the 'anchor' carrier whereas the secondary carrier is referred to as the 'non-anchor' carrier. It must be noted that, since the UE transceiver is restricted to a bandwidth of 180 kHz, it can only latch, at a time, either to the anchor or the non-anchor carrier pair.

3.2.4 Carrier deployment modes

NB-IoT can be deployed using any pair of carriers within one of the specified operating bands in Table 3-1, each with an effective carrier bandwidth of 180 kHz. As mentioned in Chapter 1, there are three deployment modes available for such a carrier pair, listed as follows:

1. In-band
2. Guard-band
3. Stand-alone

These three deployment modes are illustrated in Figure 3-2 [42]. For simplicity, only one of the NB-IoT uplink/downlink carrier pair is shown for each deployment mode. The second carrier is assumed to be deployed in a similar manner with a frequency separation.

In the *in-band* mode, the NB-IoT carrier is deployed as a carrier occupying a bandwidth (180 kHz) of a Physical Resource Block (PRB) within an LTE carrier. A PRB in LTE corresponds to the minimum schedulable time-frequency resource in the uplink and downlink (this is further illustrated in Section 3.3). This makes the in-band deployment relatively simple since most of the hardware components such as the antenna and the RF modules at the eNodeB can be reused and only a software upgrade is required. However, since the NB-IoT carrier is located within an LTE carrier, certain time-frequency resources in the downlink are reserved for LTE control channels and cell-specific reference signals which limits the maximum achievable downlink throughput for NB-IoT, compared with that in the other deployment modes. Furthermore, since the NB-IoT carrier cannot be used for LTE data traffic, it can impact the LTE capacity as well.

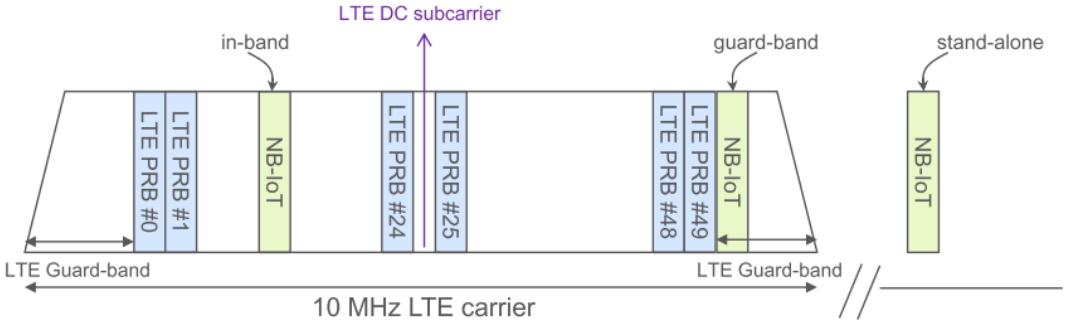

Figure 3-2. Different carrier deployment options for NB-IoT [42].

To avoid the above drawbacks with the in-band mode, the *guard-band* mode is supported, wherein the NB-IoT carrier may be deployed within the guard-band of an LTE carrier. Since it is outside the LTE system bandwidth, more time-frequency resources are available in the downlink, as compared with the in-band case. Hence, it is expected that this will result in a better downlink throughput performance as compared with the in-band case. Similar to the in-band case, the implementation can be done by a software upgrade. However, certain limitations may be imposed (e.g. by the regulator) on the amount of downlink power boosting that can be applied to the guard-band carrier so that the adjacent channel interference to the one or more neighbouring carriers, which could belong to a different operator, is minimised.

In the *stand-alone* mode, the NB-IoT carrier is deployed within a separate spectrum that is either dedicated for NB-IoT or originally reserved for non-LTE RATs such as GSM and UMTS. The latter option is known as spectrum refarming. The available number of downlink time-frequency resources for NB-IoT in the stand-alone mode is the same as that in the guard-band case. A higher base station-transmit power may also be used to further improve the downlink coverage performance. However, till date, regulators around the world have not assigned dedicated frequencies for NB-IoT [45]. Furthermore, spectrum refarming would involve both hardware and software upgrades, resulting in a higher cost of implementation. A recent study has predicted that early NB-IoT deployments are likely to adopt the in-band mode [45].

3.3 Physical layer resource structure

In this section, the physical layer resource structure of NB-IoT is discussed. First, the general time-frequency resource structure in the downlink and uplink is described. Secondly, we will look at the description and structure of the physical channels and signals defined in the downlink and uplink.

3.3.1 Time and frequency resource structure

In the following two subsections, the reader is introduced to how the NB-IoT physical layer time and frequency resources are structured, in the downlink and uplink respectively.

3.3.1.1 Downlink

Figure 3-3 [46] shows the resource grid in the downlink. Both time (horizontal axis) and frequency (vertical axis) domains are represented. A resource grid is formed by seven OFDM symbols in the time domain and twelve consecutive subcarriers with a subcarrier spacing of 15 kHz in the frequency domain. The seven OFDM symbols in the time domain form one slot, with a total duration of 0.5 ms. The smallest data-carrying element is referred to as a resource element, formed by a single subcarrier and OFDM symbol. Note that the resource grid shown in Figure 3-3 is the minimum schedulable resource in the NB-IoT downlink, just as a PRB is in the LTE downlink. However, this does not hold true for the uplink, as we shall see in the next subsection.

A set of slots are bundled further in the time domain. Figure 3-4 [46] illustrates this. We can see that two slots compose a subframe of 1 ms and ten such subframes compose a radio frame of 10 ms.

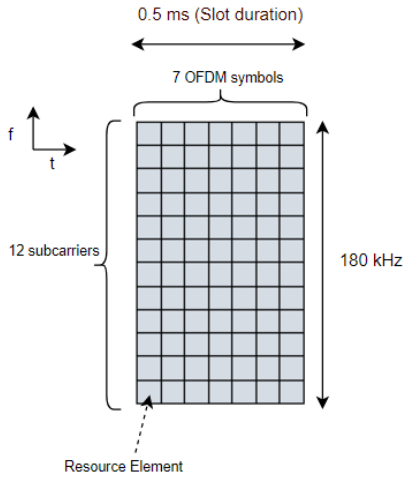

Figure 3-3. NB-IoT downlink resource grid.

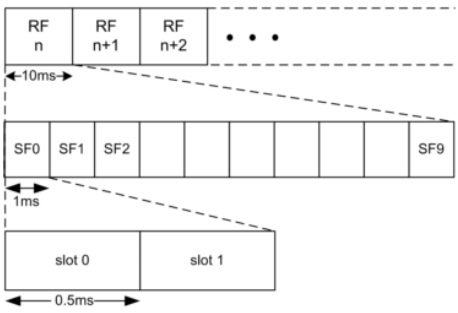

Figure 3-4. Downlink radio frame structure in NB-IoT (RF stands for radio frame, SF for subframe) [46].

As mentioned briefly in Section 1.2, to improve coverage, a key feature in NB-IoT is that of transmission repetitions of a transport block. This improves channel estimation at the receiver and increases the effective SINR of the received transport block, consequently increasing the likelihood of successful decoding. The number of possible downlink repetitions can vary from one to 2048, where the values above one are a multiple of two. An example illustration of downlink repetitions is shown in Figure 3-5. In this example, two repetitions are used. Before transmission, the transport block is mapped to four subframes. The number of subframes required for a given transport block is determined by the scheduler, based on the chosen MCS and the size (in bits) of the transport block, also referred to as the Transport Block Size (TBS). The final transmission with two repetitions is sent as a cycle, from the first subframe to the last, of two repetitions of each of the subframes, as shown in the figure.

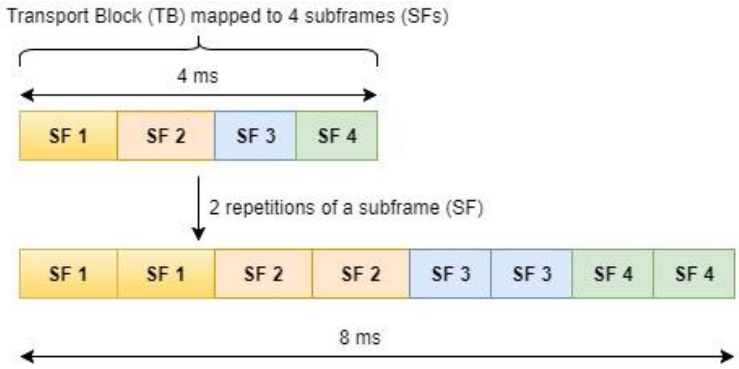

Figure 3-5. Example of downlink transmission of a transport block mapped to four subframes with two repetitions.

3.3.1.2 Uplink

The uplink resource grid for the two subcarrier spacing options (15 kHz and 3.75 kHz) is shown in Figure 3-6.

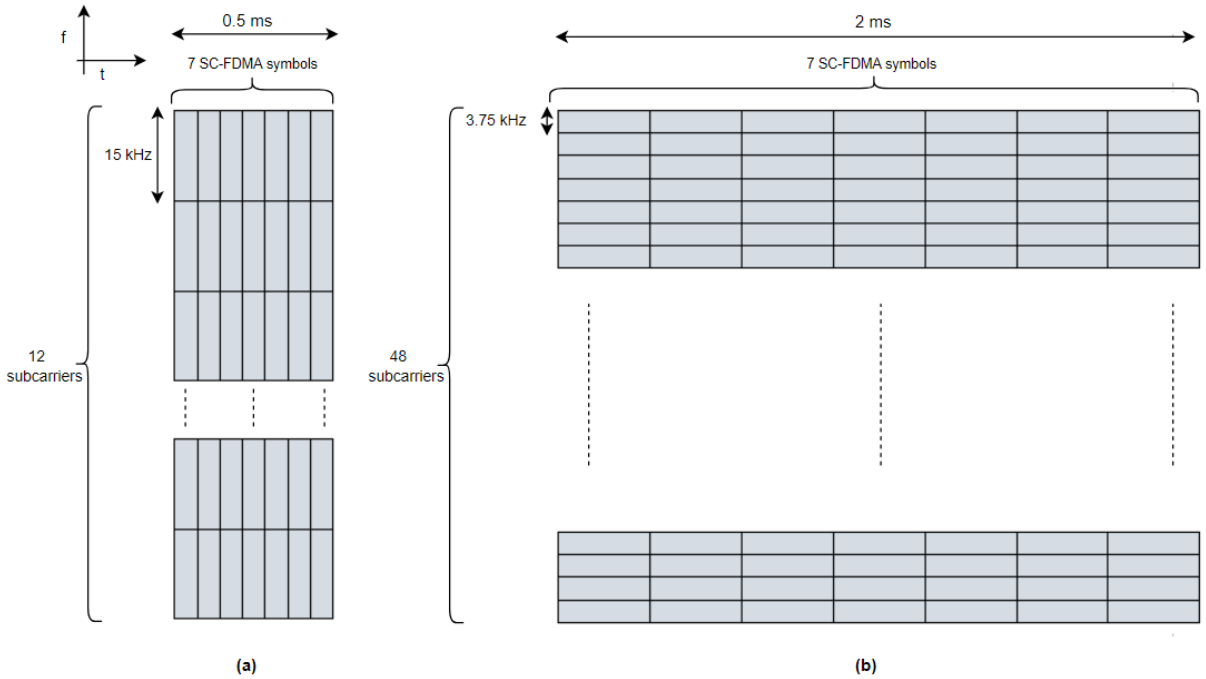

Figure 3-6. NB-IoT uplink resource grid for (a) 15 kHz and (b) 3.75 kHz subcarrier spacing.

For the 15 kHz subcarrier spacing, we can see that the slot duration (0.5 ms) in the time domain and the number of subcarriers (12) in the frequency domain are the same as that in the downlink (Figure 3-3). However, for the 3.75 kHz subcarrier spacing, the SC-FDMA symbol duration is four times larger. Hence, the duration of a slot consisting of seven SC-FDMA symbols is 2 ms, compared with 0.5 ms in the case of 15 kHz subcarrier spacing. Furthermore, the total number of subcarriers in the frequency domain is also four times

as large i.e. 48, thereby occupying the same 180 kHz as in the case of 15 kHz subcarrier spacing. The detailed slot configuration for the two subcarrier spacing options can be referred from Table 3-3.

Table 3-3. NB-IoT uplink slot configuration.

Subcarrier spacing	Number of subcarriers per carrier $N_{UL, SC}$	Number of SC-FDMA symbols per slot $N_{UL, SYM}$	Symbol duration (μs)	Slot duration T_{slot} (ms)
15 kHz	12	7	71.4	0.5
3.75 kHz	48	7	285.7	2

Schedulable resources in the NB-IoT uplink are expressed in terms of so-called 'Resource Units' (RUs), i.e., the smallest unit in the time-frequency domain to which a transport block can be mapped. The resource grid of a resource unit, in general, depends on the type of data, i.e. regular data or control information, to be transmitted, the subcarrier spacing and the number of subcarriers used. This is shown in Table 3-4.

Table 3-4. Possible combinations of number of subcarriers and slots per Resource Unit (RU).

Uplink data type	Subcarrier spacing	Subcarriers per RU $N_{RU, SC}$	Slots per RU $N_{RU, Slots}$	Duration of 1 RU (ms) $T_{RU} = T_{slot} \times N_{RU, Slots}$	Total data Resource Elements (REs) per RU[5]
Regular data	3.75 kHz	1	16	32	96
	15 kHz	1	16	8	96
		3	8	4	144
		6	4	2	144
		12	2	1	144
Control information (ACK/NACK)	3.75 kHz	1	4	8	16
	15 kHz	1	4	2	16

From the above table, we can make the following general observations:

1. With a decrease in the number of subcarriers, the length of the RU increases in the time domain. For example, with the 15 kHz subcarrier spacing, an RU with three subcarriers is four times longer compared with the length of an RU with twelve subcarriers. Thus, with an increase in the granularity, i.e. lower number of subcarriers per UE, an increase in the scheduling capacity, i.e. the number of UEs that can be served simultaneously, is achieved, at the expense of a higher duration of an RU.
2. A similar trade-off as 1. above exists between coverage and the RU duration. A lower number of subcarriers per UE increases the PSD (and hence the SINR) in the uplink, which improves coverage at the expense of a higher duration of an RU. A similar trade-off applies when choosing a lower subcarrier spacing of 3.75 kHz to improve the scheduling capacity.
3. For transmitting regular uplink data using the 15 kHz subcarrier spacing and 12 subcarriers, an RU is the same as a PRB pair in the LTE uplink, since it is two slots long and lasts 1 ms.
4. For transmitting control information, only a single-tone transmission is possible, for both subcarrier spacing options.

A single uplink scheduling grant can allow the UE to send its data using more than one resource unit, with the number of resource units a function of the amount of data required to transmit and the MCS applied. The

[5] Number of data REs = Data symbols per slot x Slots per RU. One out of seven symbols in an uplink slot is used for carrying the Demodulation Reference Signals or DMRS (explained later). Hence, only six data symbols per slot are available.

choice of MCS is dependent on the effective uplink SINR. The effective uplink SINR, in turn, depends on the number of subcarriers assigned to the UE, the subcarrier spacing and the assigned transmission repetitions per resource unit. Hence, based on the duration of an RU (determined by the data type, the subcarrier spacing and the number of subcarriers assigned), the number of repetitions and the number of resource units assigned, the overall transmission time for the UE can be determined. The probability with which the data is erroneously decoded is determined by the BLER, which is a function of the MCS and the SINR. The relationship between these multiple parameters influencing the uplink transmission time and BLER for a UE is illustrated in Figure 3-7.

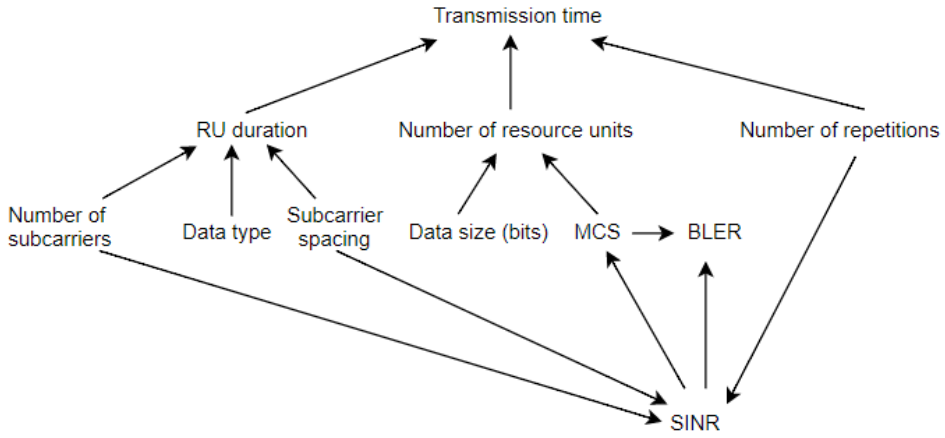

Figure 3-7. Relationship between parameters which determine the uplink transmission time and BLER for a UE.

From Figure 3-7, we can see that the choice of the subcarrier spacing, the number of subcarriers and the number of repetitions, assigned to the UE, play an important role in determining the UE's transmission time for a given data size of a certain data type. Different schemes for the choice of these parameters will be explored as part of the scheduler design.

The number of transmission repetitions in the uplink can vary from 1 to 128, where the values above 1 are multiples of 2. The repetitions of a transport block are transmitted in the uplink in a slightly different way compared with the downlink. The key difference is that the transport block is mapped to a set of resource units instead of subframes. Since a resource unit may span several slots, the repetition cycle involves only a pair of slots at a time, which is similar to the downlink case, where the repetition is of a single downlink subframe, composed of two slots, at a time. This is illustrated in Figure 3-8.

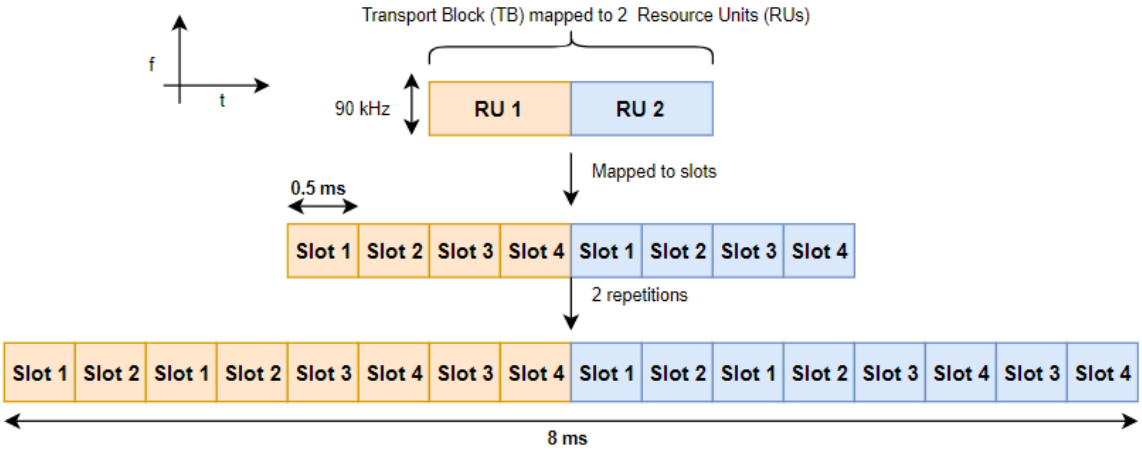

Figure 3-8. Example transmission of an uplink transport block using six subcarriers (15 kHz subcarrier spacing), two resource units and two repetitions.

The figure shows an example transmission of a transport block corresponding to regular uplink data using six subcarriers (15 kHz subcarrier spacing), two resource units and two repetitions. As per Table 3-4, each RU comprises four slots and lasts a total of 4 ms. The repetition cycle begins with the first slot pair, i.e. Slots 1 and 2 of the first RU, repeated twice and ends, similarly, with the last slot pair, i.e. Slots 3 and 4 of the second RU.

3.3.2 Physical channels and signals

In this section, we shall look at the different physical channels and signals in the downlink and the uplink, as defined in the NB-IoT standard.

3.3.2.1 Downlink

In the downlink, there are three physical channels defined as follows [44]:

1. **Narrowband Physical Broadcast Channel (NPBCH)**: used for transmission of the Master Information Block (MIB) which conveys information such as the current radio frame number, system bandwidth, and the type of deployment mode (in-band, guard-band or stand-alone) implemented in the network [46].
2. **Narrowband Physical Downlink Control Channel (NPDCCH):** carries the Downlink Control Information (DCI) which includes information about the uplink resource allocation (UL grant), ACK/NACK information for the uplink transmissions, downlink scheduling, the number of subframe repetitions for the NPDCCH and paging information [40]. There are three DCI formats defined [47], each for the delivery of different information:
    i. **DCI Format N0**: used for UL scheduling and ACK/NACK feedback on UL transmissions.
    ii. **DCI Format N1**: used for DL scheduling
    iii. **DCI Format N2**: used for conveying paging information

    For the detailed contents of each DCI format, the reader is referred to [47] and [48].
3. **Narrowband Physical Downlink Shared Channel (NPDSCH):** apart from delivering downlink data, this channel is also used for transmission of the System Information Block or SIB which contains information about the neighbour cell information for inter/intra-frequency cell reselection.

Apart from the above physical channels, there are two physical signals in the downlink:

1. **Narrowband Reference Signal (NRS):** used for cell selection and channel estimation (at the UE). For example, based on the received reference signal power levels of each cell, the UE selects the best serving cell to camp on in the idle mode.
2. **Narrowband Primary Synchronisation Signal (NPSS) and Narrowband Secondary Synchronisation Signal (NSSS):** synchronisation with the radio frame boundary is enabled by the NPSS whereas the NSSS is used to determine the cell ID. The NSSS also indicates whether the cell IDs for LTE and NB-IoT are the same or not, in the case of an in-band deployment.

Due to the limitation in the available resource elements in the frequency domain, most of the physical channels and signals are multiplexed in time rather than frequency. As depicted in Figure 3-9 [41]. There are dedicated subframes for the NPBCH, the NPSS and the NSSS whereas the NPDCCH and the NPDSCH are sent depending on the context. It will be shown in Section 3.4 how the NPDCCH is sent periodically and provides the scheduling information for the NPDSCH transmission.

The NRS (not shown in the figure) is multiplexed both in time and frequency domain with the relevant channel in a given subframe. This is done by reserving certain resource elements in time and frequency domain for the NRS, as illustrated later in Section 3.4 in the discussion of the NPDCCH resource structure. This also means that not all time-frequency resource elements in a subframe are occupied by the relevant channels/signals that are depicted in Figure 3-9. Due to space constraints, the detailed resource structure of all the relevant channels and signals is not discussed here. However, interested readers may refer to [46] for details.

	subframe number									
even numbered frame	0	1	2	3	4	5	6	7	8	9
	NPBCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPSS	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NSSS
	subframe number									
odd numbered frame	0	1	2	3	4	5	6	7	8	9
	NPBCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPSS	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH	NPDCCH or NPDSCH

Figure 3-9. Time multiplexing of physical channels and signals in the downlink [41].

3.3.2.2 Uplink

There are two uplink channels defined [44]:

1. **Narrowband Physical Random Access Channel (NPRACH)**: used for sending the preamble as part of the random access process (described in Section 3.5).
2. **Narrowband Physical Uplink Shared Channel (NPUSCH)**: used for sending regular uplink data and control information (ACK/NACK feedback for the received downlink data) as also shown in Table 3-4. These two data types are categorised into two formats, format 1 and format 2, respectively [44].

The above mentioned two uplink channels may share, fully or partially, the time-frequency resources in the uplink carrier (see Section 3.5).

In the uplink, only a single physical signal is defined, known as the **Demodulation Reference Signal (DMRS).** This is used for channel estimation by the eNodeB.

3.4 Scheduling: physical layer implementation

This section provides an overview of how downlink and uplink scheduling is implemented at the physical layer. It must be noted that the decision to schedule certain time-frequency resources to a particular UE is made by the scheduler at the Medium Access Control (MAC) layer of the eNodeB. The main aim of this section is to illustrate how a generic scheduling of time-frequency resources in both downlink and uplink is realised, at the physical layer. The knowledge of this will be used in the actual design of the scheduler presented in Chapter 5.

The reader is first introduced to the general resource structure and configuration of the NPDCCH, which carries the necessary control information for scheduling. This is followed by a short description of how time-frequency scheduling is implemented, using the NPDCCH, in the downlink and uplink.

3.4.1 NPDCCH resource structure and configuration

Figure 3-10 (modified from [46]) shows the resource structure of an NPDCCH subframe in the context of in-band deployment. The light and dark green-shaded resource elements are reserved for carrying NPDCCH related data. Note that the remaining resource elements are reserved for either the NRS or the LTE physical channels/signals due to in-band deployment. The set of NPDCCH resource elements of the same colour in the figure is referred to as a Narrowband Control Channel Element (NCCE). Thus, there are two NCCEs in an NPDCCH subframe and each NCCE spans six subcarriers in the frequency domain.

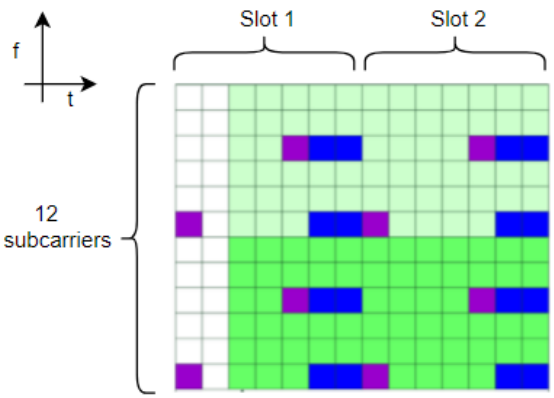

Figure 3-10. An NPDCCH subframe (in an in-band deployment) with two NCCEs (in light and dark green). The violet and blue resource elements correspond to the LTE and the NB-IoT reference signals respectively. The white-coloured resource elements are reserved for the LTE control channel known as the PDCCH.

An NPDCCH subframe carries control information corresponding to one or more categories i.e. DCI formats (mentioned in Section 3.3.2.1). The control information for a given DCI format, targeted for a specific UE, may occupy one or two NCCEs. Consequently, this also means that a given NPDCCH subframe may contain control information for either one (if both NCCEs are occupied) or two UEs (if only one NCCE is occupied). The choice of occupying one or two NCCEs is based on the target code rate of the NPDCCH transmission which is influenced by the downlink channel conditions of the UE. Further, to serve a UE in poor coverage, the NCCE may be repeated over multiple subframes in the time domain.

Since a UE cannot know beforehand when a DCI for that UE will be sent, it must listen for the NPDCCH subframes and attempt to decode them, to check if it's addressed by the NPDCCH. The NPDCCH configuration for a UE is primarily defined by how often the UE should start to listen for NPDCCH subframes and the maximum number of subframes it should listen for. These two aspects are respectively denoted by the NPDCCH period (**T**) and maximum NPDCCH repetitions ($\mathbf{R_{max}}$). This is illustrated in Figure 3-11 (a). In the figure, the NPDCCH period is assumed to start from a subframe at time **t = 0**. Thus, the UE starts to listen for the NPDCCH subframes at the following time instances: **t** = **0**, **T**, **2T**, etc. Out of the **T** subframes contained in one period, the UE listens for NPDCCH subframes for a maximum duration equal to $\mathbf{R_{max}}$ subframes, alternatively referred to as the 'non-gap interval'. The duration of the remaining $\mathbf{T - R_{max}}$ subframes, naturally, is the 'gap interval'. Figure 3-11 (a) further shows that at the start of the second period, the UE has received four repetitions of the NPDCCH, which it would then decode to check for relevant control information.

The choice of $\mathbf{R_{max}}$ (for a given period **T**) could have the following general implications:

1. A higher $\mathbf{R_{max}}$ implies a higher possible improvement in the DL SINR i.e. a better downlink coverage.
2. A higher $\mathbf{R_{max}}$ implies a higher number of opportunities to schedule the UE, in a period **T**. In high network load scenarios, this may decrease the waiting time for the UE to be scheduled. The maximum number of scheduling opportunities per second can be represented by $\mathbf{R_{max}*1000/T}$, where **T** is in ms.

The general trade-off in increasing the $\mathbf{R_{max}}$ value is that it may also increase the energy consumption of the UE since the receiver would need to be active to listen for the NPDCCH subframes for a longer time duration.

The beginning of the period may be shifted with respect to **t** = 0 using the offset parameter **α**. The number of subframes by which the period is shifted is denoted by **α** x **T** (in Figure 3-11 (a), **α** = 0). Figure 3-11 (b) shows how different **α** values could be used to create 'non-overlapping' NPDCCH configurations for three UEs having the same value of period **T** (equal to 16). Note that the term 'non-overlapping' means that the non-gap intervals of the UEs do not overlap with each other. As shown in the figure, the NPDCCH periods of UE #2 and UE #3 start two (i.e. 16/8) and six (i.e. 48/8) subframes later, respectively, compared with that of UE #1.

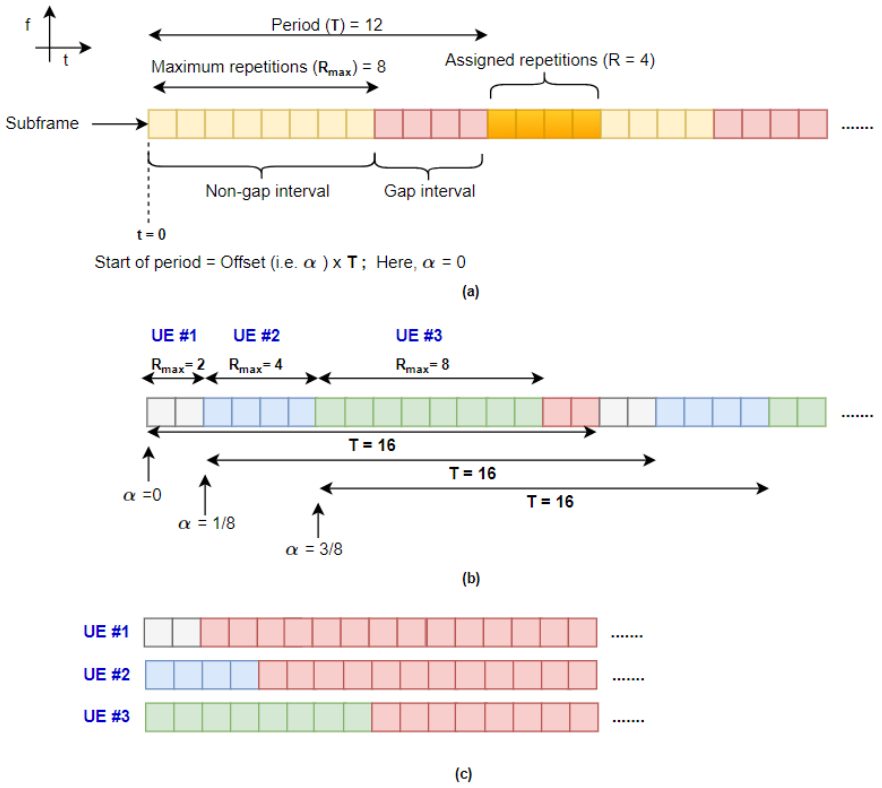

Figure 3-11. Illustration of (a) an example NPDCCH configuration for a UE, (b) the use of the offset (α) to create non-overlapping NPDCCH configurations and (c) a corresponding overlapping NPDCCH configuration.

It is useful to point out here that, in an overlapping configuration (Figure 3-11 (c)), if any of the UEs receive one or more NPDCCH subframes, it may partially or fully occupy the non-gap interval of the remaining UEs, thereby blocking the opportunity to schedule those UEs. Further, the likelihood of this blocking increases as the $\mathbf{R_{max}}$ is increased for any one of the UEs. A non-overlapping NPDCCH configuration helps to address this issue. However, the trade-off is that the minimum duration of the common period **T** is limited by the $\mathbf{R_{max}}$ value assigned to each UE. If the $\mathbf{R_{max}}$ value for a UE is increased, e.g. to provide better coverage, then **T** may have to be increased for all the UEs. A higher **T** would mean a decrease in the maximum number of scheduling opportunities per second, at least for the remaining UEs.

In principle, it may not be possible to create non-overlapping configurations for all UEs and instead, only for groups of UEs. For example, UEs may be partitioned into groups based on their experienced path loss levels. In NB-IoT, such groups are referred to as 'coverage-levels', which is further explained in Section 3.5, as part of the discussion of the NPRACH.

The possible values of relevant NPDCCH configuration parameters are listed as follows [49]:

1. **Maximum repetitions** ($\mathbf{R_{max}}$): possible values are 1,2,4… 2048
2. **Period (T)**: as per [47], the period **T** is derived using the periodicity parameter **G** where $\mathbf{T} = \mathbf{R_{max}} * \mathbf{G}$. The possible values of **G** are 1.5, 2, 4, 8, 16, 32, 48 and 64. For a given $\mathbf{R_{max}}$, **G** must be selected such that $\mathbf{T} - \mathbf{R_{max}} >= 4$ ms.
3. **Offset (α):** possible values are 0, 1/8, 1/4 and 3/8.

3.4.2 Downlink and uplink scheduling

An illustration of a generic scenario of time/frequency domain scheduling of UEs in the downlink and uplink is shown in Figure 3-12. Note that, in the time domain, boundaries of the subframes or resource units are not shown in the figure for simplicity. Hence, in general, it can be assumed that any given channel is composed of a certain number of subframes/resource units and associated number of repetitions.

In the figure, two UEs (UE #1 and UE #2) are scheduled for a time-multiplexed downlink transmission (NPDSCH) via the NPDCCH. Note that in the downlink, sub-PRB allocations are not possible. Hence, scheduling is always in the time domain. Three other UEs (UE #3, UE #4 and UE #5) are multiplexed in the time and frequency domain for uplink transmission (NPUSCH) with the uplink resource assignments signalled via the NPDCCH. The frequency-domain multiplexing in the uplink is shown by simultaneously serving, for a certain duration, two UEs (UE #3 and UE #4) with six subcarriers each. UE #2 and UE #3 have been scheduled simultaneously by the same set of NPDCCH subframes, which is achieved by using one NCCE, for each UE, to carry the respective control information.

As shown in the figure, the time between the end of the NPDCCH subframe and the start of a scheduled DL or UL transmission is denoted by a so-called 'scheduling delay'. This serves three main purposes:

1. It allows the UE the time to decode the control information in the NPDCCH.
2. It brings flexibility in the scheduling and consequently reduces power consumption in the UE. For example, in Figure 3-12, in the uplink, the scheduler exploits the scheduling delay to schedule UE #5 in advance even though the resources were fully occupied at the end of the NPDCCH subframe. Once UE #5 is scheduled, it need not listen for further NPDCCH subframes until the completion of the uplink transmission, and the receiver can go to a standby mode which reduces the energy consumption.
3. It accounts for the guard period required in a half-duplex operation. Note that this is only needed during UL scheduling to switch from the reception mode, after receiving the NPDCCH message, to the transmission mode, to send the UL data.

As per [47], only certain discrete values of scheduling delay are possible, between 8 and 64 ms for the UL and between 4 and 1028 ms for the DL.

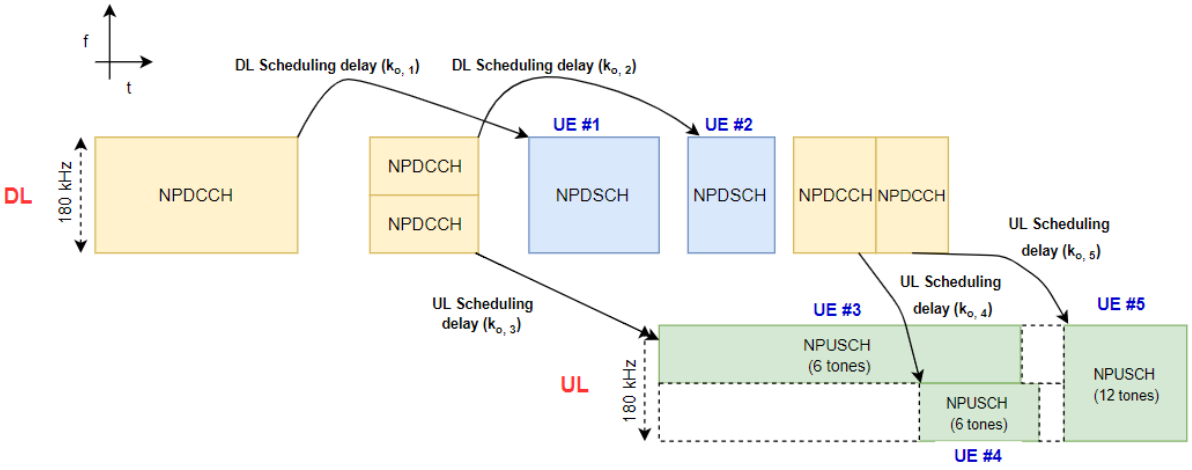

Figure 3-12. Illustration of a generic scenario of DL (top) and UL (bottom) scheduling of UEs in NB-IoT.

3.5 Random access and small data transmission in NB-IoT

We saw in Chapter 2 that data transmission in smart meters is either periodic or on-demand/event-based. In both cases, the inter-arrival time of packets for a UE was very high, e.g. in the order of hours or days. For reduced power consumption between successive transmissions, the NB-IoT device would typically transit to an 'idle' state where it only performs minimal necessary functions such as listening to paging/system information messages, performing cell (re) selection, and does not listen to any control information (NPDCCH). If the UE needs to send uplink data, it must notify the network and wait for resources to be allocated for data transmission. This is achieved using the contention-based random access process which is essentially an exchange of physical and higher layer signalling information between the UE and eNodeB. The regular LTE random access process involves the exchange of several signalling messages before the actual data transmission starts. This would result in a large overhead for NB-IoT devices which typically send small

infrequent payloads. In this section, we will look at the random access procedure in NB-IoT (including the role of the random access channel, i.e. the NPRACH) and how it is slightly optimised to reduce the signalling overhead. First, an overview of the key steps in the random access procedure is provided, including how the signalling overhead is reduced compared to regular LTE. Then, the resource structure and configuration of the NPRACH is described.

3.5.1 Key steps

The overall random access procedure call flow is depicted in Figure 3-13. It starts with one or more UEs in an 'idle' state independently selecting a physical layer signature, called preamble or MSG 1, to send on the random access channel (NPRACH). The available set of preambles are designed such that they are orthogonal to each other, as will be shown in the next section. Hence, if there are N such preambles, then simultaneous preamble transmissions by upto N UEs can in principle be decoded successfully. Since the preambles are selected independently by a UE, it is however possible that multiple UEs may select and send the same physical layer preamble, which can result in a 'collision' and an unsuccessful decoding of the preamble at the receiver (eNodeB). Note that, in the event of a collision, the eNodeB may still be able to decode a preamble successfully if one of the preambles has sufficiently higher received signal power than the rest (known as the 'capture' effect). If the preamble is successfully decoded at the eNodeB, the UE is sent MSG 2, called the Random Access Response (RAR), scheduled on the NPDSCH. If the scheduling message, on the NPDCCH, for the RAR is not received within a predefined time window, the UE reattempts the preamble transmission after a randomly chosen backoff time. Upon decoding the RAR, the UE obtains the uplink grant for MSG 3.

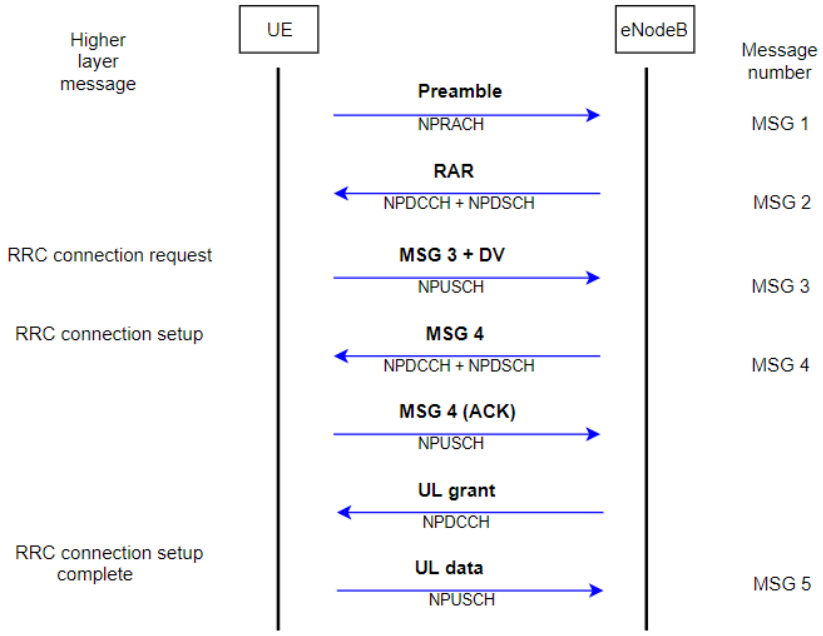

Figure 3-13. Random access procedure call flow in NB-IoT. The corresponding higher layer message is indicated on the left and the message number indicated on the right.

Once the uplink grant for MSG 3 is received and decoded successfully, the UE sends MSG 3, containing a UE-specific identifier, a higher layer message (RRC connection request) and a so-called 'Data Volume (DV).' The purpose of the DV, which is a new feature added to the legacy LTE random access process, is to roughly indicate to the eNodeB the amount of data present in the UE buffer. The DV is sent as an index value (between 0 and 15) that corresponds to a range of data size in bytes, as shown in Table 3-5 [50].

Table 3-5. Data Volume (DV) index and corresponding range of data size (bytes) [50].

Index	Data Volume (DV) value [bytes]	Index	Data Volume (DV) value [bytes]
0	DV = 0	8	67 < DV <= 91
1	0 < DV <= 10	9	91 < DV <= 125
2	10 < DV <= 14	10	125 < DV <= 171
3	14 < DV <= 19	11	171 < DV <= 234
4	19 < DV <= 26	12	234 < DV <= 321
5	26 < DV <= 36	13	321 < DV <= 768
6	36 < DV <= 49	14	768 < DV <= 1500
7	49 < DV <= 67	15	DV > 1500

Since the RAR is addressed to each successfully received preamble, UEs which selected the same preamble would receive the same MSG 3 grant, conveyed in MSG 2, resulting in a possible second uplink collision, i.e. of the MSG 3 transmissions. If one of the colliding MSG 3s is successfully decoded, the UE is identified through its UE-specific identifier (part of MSG 3) and a 'contention resolution' message MSG 4 is sent, i.e. the RRC connection setup message, containing this UE-specific identifier. Similar to the RAR, there is also an associated time window for the reception of the NPDCCH associated with MSG 4. MSG 4 is followed by a UL grant to enable the UE to send MSG 5 (RRC Connection Setup Complete). Based on the DV sent in MSG 3, in the UL grant, the eNodeB allocates some additional uplink resources for the UE to also allow it to transmit its small payload (UL data in the call flow) in MSG 5. In regular LTE, this feature is not supported and there are at least four more signalling messages exchanged before the actual uplink data transmission takes place. Thus, with this feature, small payloads can be sent in NB-IoT with reduced delay and signalling overhead.

It is possible that the UL grant before MSG 5 is unable to allocate sufficient uplink resources for the UE to transmit all its data within MSG 5, e.g. if the data size is too high or if the UL SINR is too low. In such cases, the remaining data can be sent after MSG 5, as part of one or more higher layer (NAS[6]) messages known as 'UL information transfer' [46].

3.5.2 Random access channel (NPRACH)

The NPRACH plays a key role in the first step, i.e. the preamble transmission, of the random access procedure described in the previous section. The NPRACH basically provides the time-frequency resources for sending the preamble.

A random access preamble in NB-IoT is a physical layer waveform consisting of modulated Zadoff-Chu sequences transmitted via a frequency hopping scheme with a 3.75 kHz subcarrier spacing single-tone transmission [44]. The frequency hopping scheme is introduced to facilitate uplink timing estimation at the eNodeB [23].

In the time-domain, the preamble consists of a set of four so-called 'symbol groups.' Each symbol group consists of a cyclic prefix (CP) and five identical[7] OFDM symbols. The combined duration of the five symbols is a fixed value, which is roughly 1.3 ms. The CP duration, on the other hand, has two values depending on the type of CP: a normal CP (66.7 µs duration) and an extended CP (266.7 µs duration). These two CP types exist to handle the impact of propagation delays in relatively small and large cell sizes respectively. Thus, the duration of a single symbol group, i.e. a CP + five OFDM symbols, is roughly 1.4 ms (with a normal CP) or 1.6 ms (with an extended CP). From this, the effective duration of a single preamble, i.e. four symbol groups, is either 5.6 ms (with a normal CP) or 6.4 ms (with an extended CP).

[6] NAS or Non-Access-Stratum is the topmost layer in the protocol stack.

[7] The use of identical symbols with a common CP is to reduce the relative time overhead of a CP, which may be of a long duration, considering that a NB-IoT deployment may target large cell sizes e.g. between 30 – 40 kms, resulting in a high propagation delay) [23].

Each preamble is identified by its frequency hopping sequence which in turn is dependent on the starting subcarrier [46]. Hence, there are as many unique frequency hopping sequences and preambles as there are starting subcarriers. If two UEs randomly select the same starting subcarrier at the same time, their hopping sequences overlap and hence their preambles may collide.

To overcome the effects of the propagation path loss that the preamble transmission may experience before reaching the eNodeB, a single preamble may be repeated a certain number of times (between two and 128) depending on the specific network configuration (explained later in this section).

An example of a preamble transmission with two configured repetitions is shown in Figure 3-14 (modified from [51]). In this figure, there are twelve starting subcarriers (preambles) and consequently twelve frequency hopping sequences. Each of the twelve sequences is represented by a different colour. We can see that a single preamble repetition comprises four symbol groups. Note that the term 'repetition,' as also used in the standard, is slightly misleading since the frequency hopping sequence does not actually repeat. This is because the frequency hopping sequence of the symbol groups after the first repetition is determined by using the repetition number itself as one of the inputs.

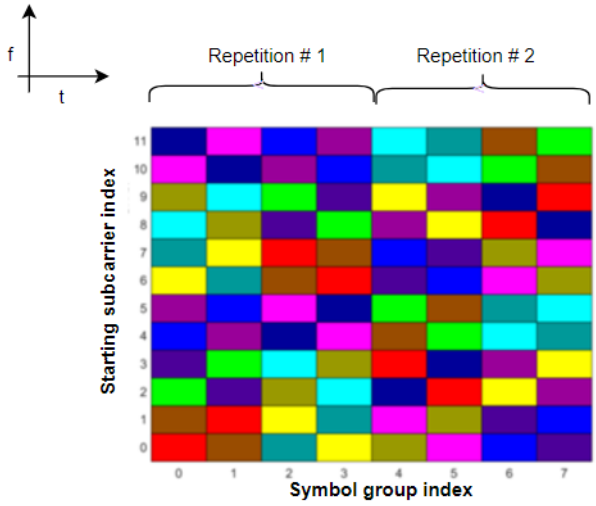

Figure 3-14. Example NPRACH preamble transmission. There are twelve starting subcarriers i.e. preambles, and each colour represents a different frequency hopping sequence.

As stated earlier, repetitions in the preamble are configured to overcome path loss effects. In order to serve UEs that experience different levels of path loss and require a different number of repetitions, the UEs are split into groups based on their so-called 'coverage levels'. Upto three coverage levels can be configured in the network, where each coverage level is defined by a Reference Signal Received Power (RSRP) threshold. The higher the coverage level, the lower the corresponding RSRP threshold, i.e. the higher the correspondingly allowed path loss. Each UE measures the cell RSRP and selects the lowest possible coverage level based on the defined thresholds. Figure 3-15 shows a simplified spatial representation of three coverage levels around the base station. Note that in this simplified representation, shadowing and penetration loss effects are not taken into account. In the figure, we can see that the lowest coverage level (CL 1) represents UEs close to the base station, and as their distance from the base station increases, resulting in a higher path loss, they are grouped into higher coverage levels. Separate non-overlapping NPRACH resources with a different predefined number of preamble repetitions are then configured for each coverage level. Consequently, the random access contention is limited to only the UEs in the same coverage level.

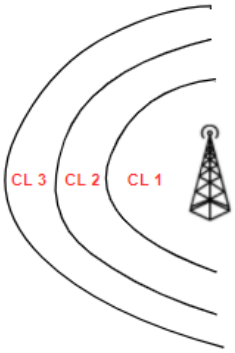

Figure 3-15. Simplified spatial representation of different coverage levels around the base station (CL stands for coverage level).

The NPRACH configuration per coverage level is defined by the following parameters:

1. **Period**: defines how often a UE in each coverage level can attempt to send a random access preamble. Within a period, there is only one opportunity (per UE) to send a preamble. Possible values are 40 ms, 80 ms, 160 ms, 240 ms, 320 ms, 640 ms, 1280 ms and 2560 ms.
2. **Number of preamble repetitions**: defines the number of repetitions of a single preamble. Possible values are 1, 2, 4, ..., 128.
3. **Number of starting subcarriers**: defines the width of the NPRACH resource in the frequency domain. The larger the number of starting subcarriers, the larger is the number of available preambles. Possible values are 12, 24, 36 or 48.
4. **Starting subframe of NPRACH opportunity**: defines the subframe after the start of the period when the preamble can be sent. Possible values are 8 ms, 16 ms, 32 ms, …,1024 ms.
5. **Location of the starting subcarrier**: defines the lowest subcarrier index at which the configured resource starts. Possible values are 0, 12, 24, 36, 2, 18 and 34.
6. **Maximum preamble attempts per coverage level**: defines the maximum preamble attempts per coverage level. Once this limit is exceeded, the UE selects the next higher coverage level if available, otherwise, declares a random access failure. The minimum attempts required are three.
7. **RAR window size**: defines the duration of the RAR time window in terms of the maximum number of NPDCCH periods (**T**). Possible values are 2, 3, 4, 5, 6, 7, 8 and 10.
8. **MSG 4 window size**: defines the duration of the MSG 4 time window in terms of the maximum number of NPDCCH periods (**T**). Possible values are 1, 2, 3, 4, 8, 16, 32 and 64.

Two example NPRACH resource configurations (A and B) are shown in Figure 3-16. Note that the resource configuration only includes the first five parameters above. In Configuration A, all 48 subcarriers are allocated to each coverage level. Hence, the NPRACH frequency resources for each coverage level must be separated in time by a single common period. In Configuration B, the resources are frequency-multiplexed, with different periodicities to each coverage level. The periodicities, in general, are assigned based on the expected usage intensity for each coverage level and the number of repetitions of the preamble, which in turn depends on the RSRP thresholds defined for each coverage level. In Configuration A, the minimum overall periodicity is limited by the combined time durations of the three preambles, whereas in configuration B the periodicity can be set independently for each coverage level. However, in Configuration B, there is some limitation in the scheduling flexibility of NPUSCH resources since the availability of all 48 subcarriers to choose from, for scheduling NPUSCH transmissions, is relatively lesser since the frequency resources are often occupied partially by the NPRACH.

Since an NPUSCH transmission can be of a long duration due to repetitions, it is possible that it may overlap with the reserved NPRACH resource. In that case, the NPUSCH transmission is postponed to the next earliest uplink slot following the end of the NPRACH time resource [44]. As result, the choice of allocating a certain amount of NPRACH resources can have a significant impact on the performance, with regards to collision probability and transmission delay, when the network load is high, e.g. if many smart meter UEs attempt network access after a power outage. This is explained below.

In general, the allocation of random access channel resources is expressed in terms of the so-called 'Random Access Opportunities (RAOs) per second.' The number of RAOs per second is equal to the number of preambles divided over the random access channel period (see configuration parameters above). For a Poisson arrival process with arrival rate $\boldsymbol{\lambda_T}$ and **L** RAOs per second, the preamble collision probability ($\mathbf{P_c}$) is given by the below equation [52].

$$P_c = 1 - e^{-\frac{\lambda_T}{L}} \tag{3.1}$$

Thus, for a given arrival rate of traffic, an increase in the number of RAOs per second decreases the preamble collision probability. If the number of RAOs per second is too high, e.g. due to a low NPRACH period value, then it may help to minimise the collision probability. However, due to lack of sufficient NPUSCH resources, the UL transmission delay may increase. Conversely, if the number of RAOs per second is too low, it may help to minimise the UL transmission delay at the expense of a high collision probability. Thus, an optimal NPRACH resource configuration may be required that balances the impact on both the collision probability and UL transmission delay.

The impact of different NPRACH resource configurations on the performance, under different network loads, is later presented in Chapter 6.

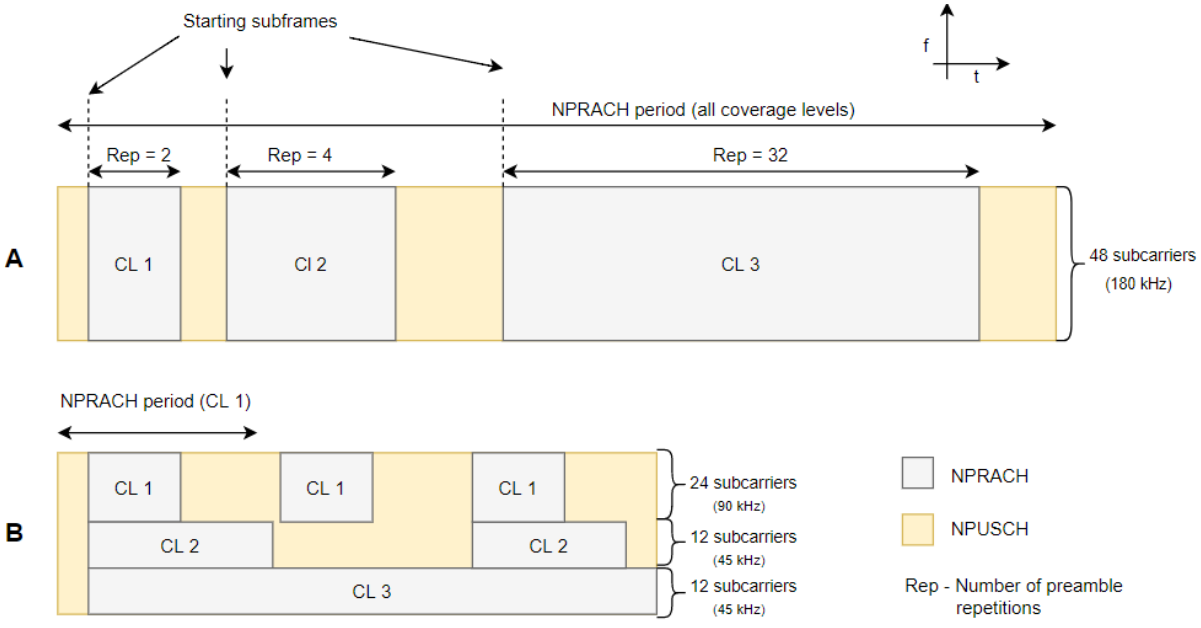

Figure 3-16. Example NPRACH resource configurations (A and B) for three Coverage Levels (CLs).

3.6 Summary

In this chapter, the reader was provided with an introduction to the different elements of the NB-IoT technology, that form a basis for the simulation model and scheduler design.

First, we looked at how the air interface has been designed to address the needs of massive IoT devices and application characteristics. The air interface is designed specifically to support low device complexity/cost, reduced power consumption and a high number of connected devices. This is done using a narrow bandwidth of 180 kHz, flexible UE uplink transmission bandwidths (single-tone and multi-tone), a reduced subcarrier spacing option in the uplink (3.75 kHz), half-duplex FDD operation and transmission repetitions. Additionally, we saw that there are three carrier-deployment modes available for an NB-IoT operator: in-band, guard-band and stand-alone. Both in-band and guard-band modes of deployment offer the advantages of easy implementation (through software upgrades) in an existing LTE site. However, the in-band mode suffers from lower downlink throughput performance, as compared with the guard-band mode, due to the overhead of LTE resource elements in the downlink system bandwidth. At the same time, there may be constraints on the level of downlink power boosting that can be applied on the NB-IoT carrier in the guard-band mode, to control the adjacent channel interference on the neighbouring carrier. The stand-alone mode does address the issues of both the in-band and guard-band modes. However, it suffers from the disadvantages of higher implementation cost if spectrum refarming is done and the lack of otherwise dedicated spectrum.

Second, an overview of the physical layer resource structure was provided. The downlink and uplink time-frequency resource structure and the NB-IoT physical channels were presented. We observed that the downlink transmission is based on OFDMA with a 15 kHz subcarrier spacing which results in a similar frequency resource grid and radio frame structure as in LTE. However, due to the narrower bandwidth, most of the physical channels and signals are multiplexed in time. In the uplink, although the transmission scheme is SC-FDMA as used in the LTE uplink, a new concept of 'resource unit' is introduced, to associate with the single and multi-tone scheduling feature. Further, to improve coverage, uplink/downlink data and control channels are repeated in time.

Third, an overview of the physical layer implementation of scheduling was presented. We looked at the resource structure of the control channel i.e. the NPDCCH, which is used to deliver the relevant control information for scheduling. An operator may configure the NPDCCH resource using three key parameters: the maximum NPDCCH repetitions ($\mathbf{R_{max}}$), period (**T)** and offset ($\boldsymbol{\alpha}$). A higher $\mathbf{R_{max}}$ allows for higher improvement in the downlink coverage and higher opportunities to schedule a UE. However, the trade-off is a higher power consumption for the UE. Further, in an overlapping configuration, an increase in $\mathbf{R_{max}}$ for one UE increases the likelihood of blocking of NPDCCH resources of other UEs. This can be addressed by configuring non-overlapping NPDCCH resources, with a common **T** value, for groups of UEs. However, the trade-off is that **T** may need to be set at a higher value, thus decreasing the maximum number of scheduling opportunities per second. Downlink scheduling in NB-IoT is only in the time domain since sub-PRB allocations are not allowed, whereas, in the uplink, time-frequency domain scheduling is possible.

Finally, an overview of the NB-IoT random access process and the NPRACH were presented. To minimise the signalling overhead in transmitting small payloads, the NB-IoT random access call flow has been slightly optimised compared to the one in legacy LTE. This is done by 'piggybacking' the data over signalling resources. To serve UEs that experience different levels of path loss, they are mapped to one of three coverage levels (based on RSRP thresholds), each of which can have a dedicated NPRACH resource configuration. Since NPRACH and NPUSCH share the same frequency resources (partially or fully), the choice of the NPRACH resource configuration may impact the preamble collision probability and the transmission delay performance. An increase (decrease) in the RAOs per second may decrease (increase) the collision probability, with the trade-off involving a higher (lower) transmission delay, due to a decrease (increase) in the NPUSCH resources.

Chapter 4. Simulation model

4.1 Introduction

As stated in Chapter 1, one of the steps in the research approach is the development of a simulation model for the performance evaluation of NB-IoT in the context of ORM. This simulation model incorporates the relevant specifications of the NB-IoT standard, as discussed in Chapter 3, the relevant characteristics of the energy distribution network of a power grid, the resource scheduler and other components which are typically part of system-level simulations of mobile networks.

The outline of this chapter is as follows. In Section 4.2, a high-level overview of the different components of the simulation model and how they interact with each other is provided. In Section 4.3, a more detailed discussion of each of the model components is done. Finally, a summary is provided in Section 4.4.

4.2 Overview of the simulation model and components

Broadly, the simulation model consists of the following six components:

1. **Network layout**, which models the radio network and energy distribution network layout;
2. **Traffic model**, which models the spatio-temporal traffic arrival process and payload sizes;
3. **Wireless channel propagation model**, which specifies the coupling loss between the UE and the eNodeB. The coupling loss between a UE and eNodeB includes aspects such as the path loss, antenna gains and shadowing loss;
4. **Physical (PHY) layer model**, which abstracts the physical layer link of the communication network, via performance parameters such as the BLER in relation to the SINR at the receiver;
5. **Media Access Control (MAC) layer model,** which models key MAC layer functions such as scheduling and link adaptation at the eNodeB. Note that in NB-IoT, link adaptation decisions for both downlink and uplink transmission are done at the MAC layer of the eNodeB;
6. **Simulation flow**, which models the state of the system and its change over time, e.g. modelling the call flow of the random access procedure (see Figure 3-13) for a set of UEs;

The simulation process comprises of a series of so-called ‘snapshots’, for a given combination of simulation inputs. These inputs typically define a particular network configuration (e.g. NPRACH) and deployment scenario (e.g. urban environment and in-band deployment) whose performance is to be evaluated. At the start of each snapshot, the network layout and/or the spatio-temporal traffic arrival process are initialised in a random manner. Note that for every six (roughly) snapshots, the network layout is fixed and only the traffic model is randomly initialised. The random initialisation of the network layout involves the positioning of the energy distribution network layout with respect to the radio network, as described further in Section 4.3.2. The random initialisation of the spatio-temporal traffic arrival process (see Section 4.3.4) involves the positioning of UEs in the coordinate axes and determining the arrival times of packets to be sent, each based on a certain probability distribution. During a snapshot, which begins with the first packet arrival and ends with the last successful transmission, the six model components interact with each other, in parallel. In the end, KPIs (e.g. reliability, success rate) for that snapshot are generated. After 100 such snapshots, the final KPIs are statistically derived from the snapshot-specific KPIs. This overall process is illustrated in Figure 4-1.

The interaction between the above six components (encircled in the figure) is explained at a high-level as follows:

1. Based on the environment (e.g. urban, rural) input, the *network layout* specifies the **coordinates** of the relevant components (e.g. UEs, eNodeBs, substations) of the radio and energy distribution networks.

2. The *wireless channel propagation model* uses the coordinates of UEs and eNodeBs, and the environment input, to determine the downlink or uplink **coupling loss**.
3. The *PHY layer model* computes the downlink or uplink SINR at the receiver. This is done based on inputs such as the UE/eNodeB transmit power, the coupling loss between the relevant UE-eNodeB pair, the coupling loss between the interfering UE-eNodeB pairs, the number of transport block **repetitions**, the number of **subcarriers**, and noise parameters such as the thermal noise and receiver noise figure. The PHY layer model, additionally, abstracts the link performance using the BLER-SINR curve for the assigned **MCS**. Based on the computed BLER, a biased coin is flipped, the decoding of the received transport block is flagged as successful or unsuccessful and accordingly, an **ACK** or a **NACK** is fed back to a link adaptation loop at the MAC layer.
4. At the *MAC layer* of the eNodeB, the link adaptation loop, based on the **ACK/NACK** feedback, updates the choice of MCS/repetition for the next (re)transmission. Amongst the UEs waiting to receive downlink data or transmit uplink data, the **scheduler** at the MAC layer decides the priority order in which to serve the UEs and the time-frequency resources to be allocated to the selected UE. The scheduler may need additional inputs such as the NPDCCH configuration to determine the set of UEs which can be scheduled at a given time instant.
5. The *simulation flow* basically models the variation of the system with time, considering the '**state**' (see Section 4.3.1) of all UEs, the **traffic arrival** process specified by the *traffic model*, outputs from the scheduler and system-level overheads such as the guard periods needed for half-duplex operation.

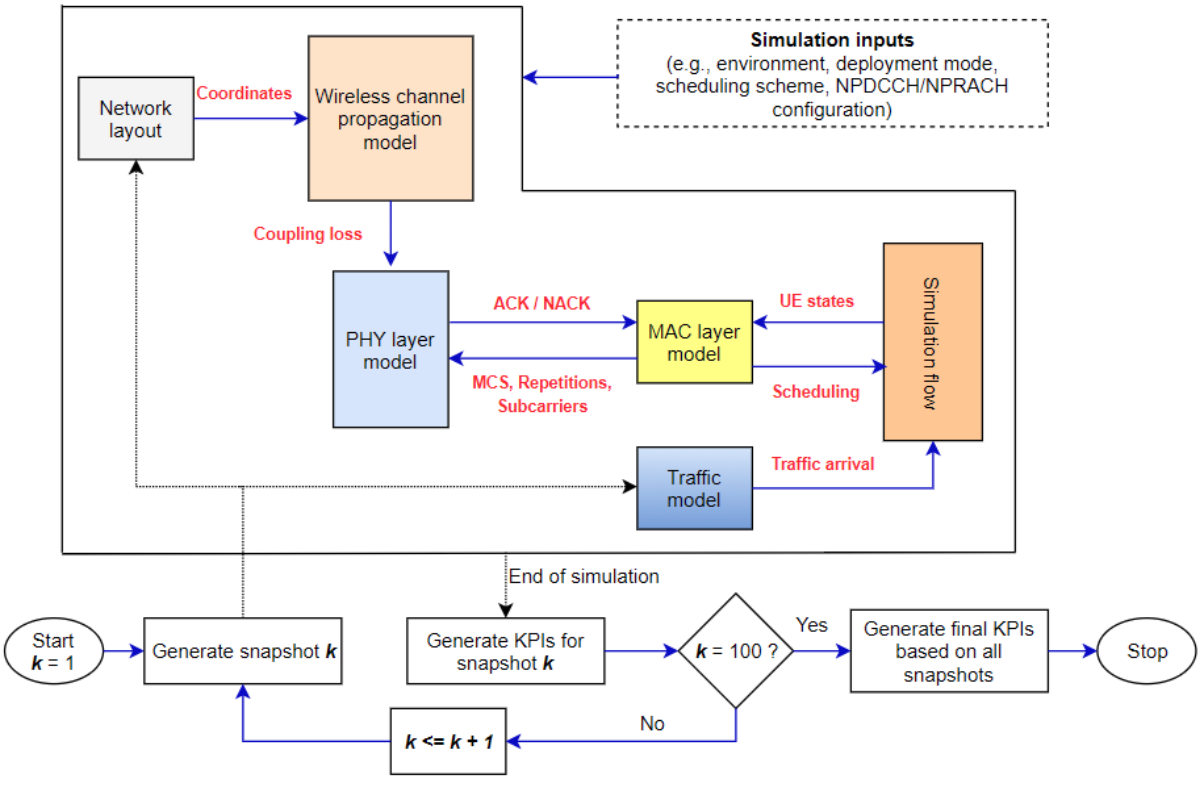

Figure 4-1. Overview of simulation process and the interaction between the six components (encircled) of the NB-IoT simulation model.

The implementation of the above simulation model is done in MATLAB [53] due to its easy-to-implement vector and multi-core processing functions, which helps to minimise the running time of large simulations.

4.3 Description of the model components

In this section, the six model components in Figure 4-1 are discussed in the following order:

1. Simulation flow
2. Network layout
3. Wireless channel propagation model
4. Traffic model
5. PHY layer model
6. MAC layer model (link adaptation)

Note that in this chapter, only link adaptation is discussed as part of the MAC layer model. Scheduling, another key aspect of the MAC layer, is discussed separately in Chapter 5.

4.3.1 Simulation flow

The general simulation approach in each snapshot is discrete event-based in which a discrete set of events (described later in this section) are modelled in time. Between two consecutive events, the state (described later in this section) of the system, represented by a set of UEs, does not change. If the time of occurrence of an event is known beforehand, then the simulation flow can skip to that event. Such an approach helps to reduce the running time of the simulation compared to a continuous time approach.

The overall simulation flow for each simulation snapshot is shown in Figure 4-2. The first four phases in the flow represent the initialisation of the simulation. To reduce the simulation complexity, only a single radio network cell and its associated UEs are considered for simulation. However, the existence of neighbouring cells is not completely ignored since their impact is modelled in the cell selection and interference calculation process. This is explained later in Sections 4.3.3 and 4.3.5 respectively.

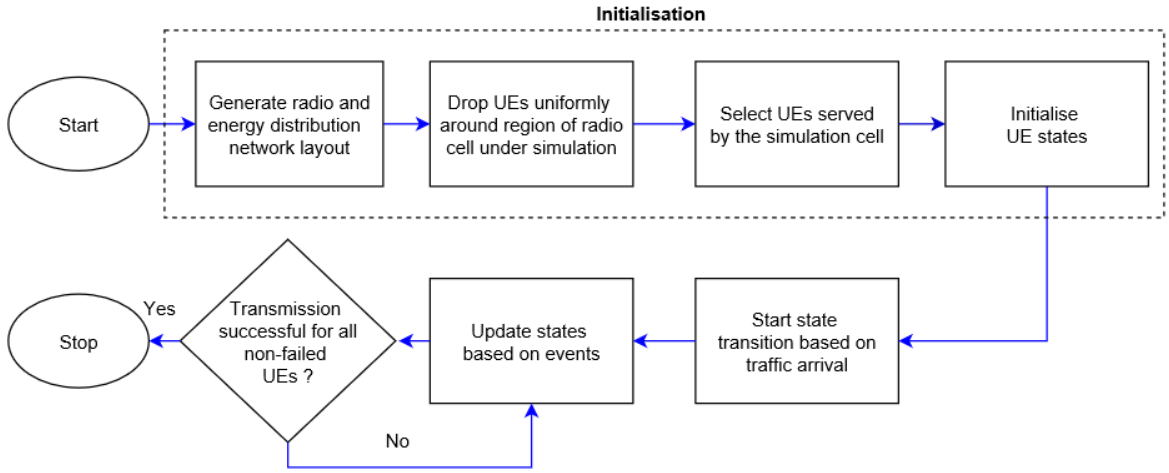

Figure 4-2. High-level simulation flow per simulation snapshot.

The high-level steps in the simulation flow are described as follows:

1. The radio and energy distribution network layout is generated on the coordinate axes. This includes mainly the placement of radio cell sites and distribution substations based on the typical characteristics of an environment such as a rural or an urban area.
2. UEs are dropped uniformly around the region of the radio cell under simulation.
3. Using the wireless channel propagation models and the UE locations, the UEs served by the cell under simulation are selected.
4. For the selected UEs, their states are initialised to 'idle' i.e. there is no packet to send in the buffer.
5. As soon as any UE gets 'activated' i.e. a packet has arrived in its buffer, a state transition is initiated which essentially represents the different stages the UE may go through in the process of sending its

data. In the context of this study, this refers to the random access procedure, followed by the small data transmission (discussed in Section 3.5).

6. Using a set of decision rules, different events for each activated UE are checked and accordingly their states are updated.
7. The simulation ends once all non-failed UEs have completed their transmission.

In the description of the high-level simulation flow, we saw that every UE undergoes a state transition based on certain events, once it is activated. This event-driven state transition process is shown in Figure 4-3.

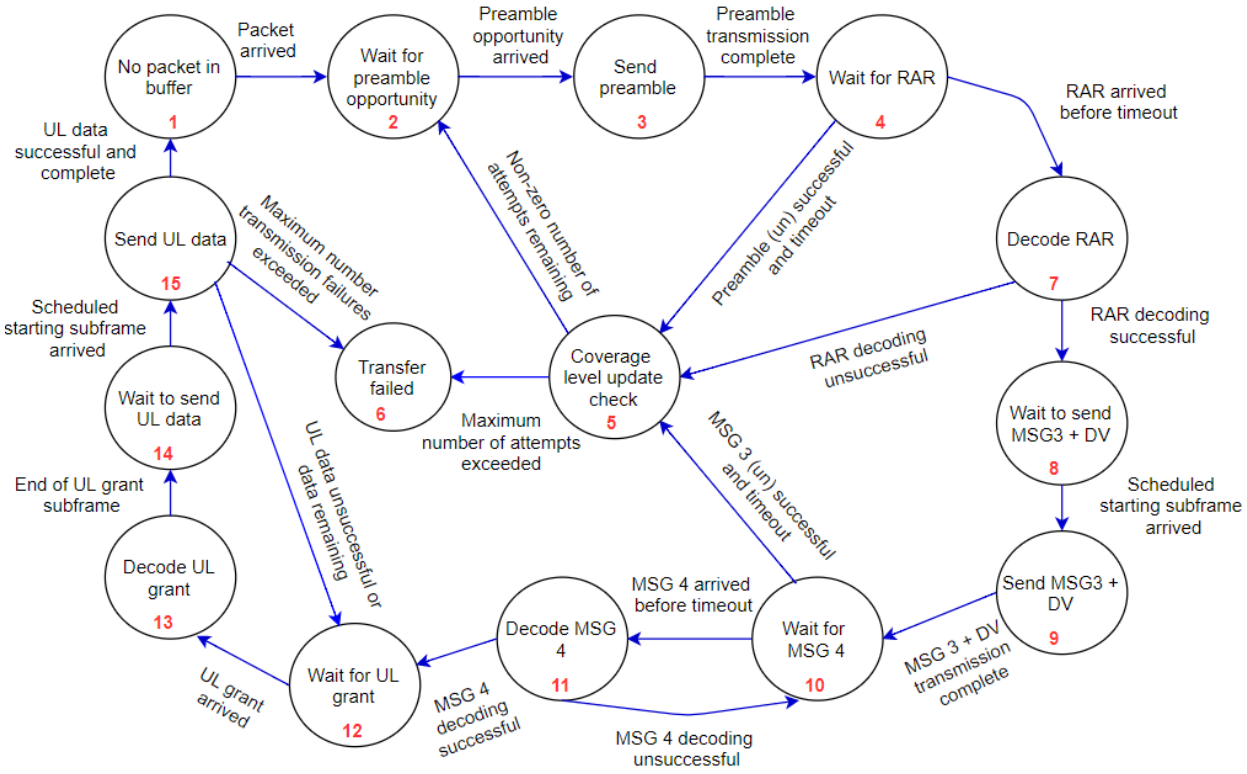

Figure 4-3. Event-driven state transitions for a UE attempting to send uplink data.

There are 15 states in total which are explained as follows:

1. In *State 1*, the UE has no packet in its buffer. In this initial state, the UE also determines its coverage level based on the measured RSRP and comparing it with the predefined coverage level thresholds. Once the packet arrives, the UE initiates the random access process.
2. In *State 2*, the UE waits for the preamble transmission opportunity i.e. the starting subframe, of its coverage level-specific NPRACH resource. Once the preamble transmission opportunity has arrived, it begins the preamble transmission.
3. In *State 3*, once the preamble transmission is complete, the UE transitions to *State 4* where it waits for the RAR in the downlink.
4. It is assumed that the preamble detection is successful only if no collision occurs and the preamble received power is above a defined threshold. If the RAR is not received within the RAR window, regardless of whether the preamble detection is successful or not, the UE enters the 'coverage level update check' state *(State 5*).
5. Within *State 5*, it is determined whether the UE can reattempt the preamble transmission or not. In general, every new preamble reattempt occurs after a randomly chosen backoff time which is uniformly sampled from a configured interval for the coverage level. Typically, a certain maximum number of random access attempts are configured per coverage level. Once the maximum attempts are reached and the UE fails, it updates its coverage level to the next higher one.

6. If the UE is already in the highest coverage level and has exhausted the maximum number of attempts, a transfer failure is declared (*State 6*).
7. Once the RAR is received within the time window, the UE attempts to decode it. If unsuccessful, it transitions to *State 5* and reattempts random access, if possible. If the RAR is decoded successfully, it transitions to *State 8* where it waits to send MSG 3, along with the data volume i.e. DV.
8. Once the scheduled starting subframe of the MSG 3 transmission has arrived, the UE begins the MSG 3 transmission in *State 9*.
9. Once MSG 3 is sent, it transitions to *State 10* where it waits to receive MSG 4 (i.e. contention resolution) within the MSG 4 time window.
10. If the UE fails to receive MSG 4 within the window, it reattempts random access (*State 5*), if possible. If MSG 4 is received within the expected time window, the UE attempts to decode it, in *State 11*.
11. If MSG 4 decoding is unsuccessful, a retransmission is requested via a NACK response, the UE returns to *State 10* and the window timer is restarted. Following a successful MSG 4 decoding, the UE awaits a further UL grant to send the UL data until all the data has been successfully sent. This is done between *States 12 to 15* which are similar to *States 7 to 9*. Additionally, as shown in the transition from *State 15* to *State 6*, a transfer failure is also declared if the UL data transmission fails for more than a threshold number of times.

The detailed decision rules for the state transition can be referred from Appendix A.

4.3.2 Network layout

The network layout model comprises the *radio network layout* and the *energy distribution network layout*, described in the following subsections.

4.3.2.1 Radio network layout

The radio network is modelled as a hexagonal grid of cells shown in Figure 4-4. Each hexagon roughly models the coverage area of a cell. The actual radio coverage area, in practice, is rather irregular, as we will see later, due to propagation effects. A total of 19 cell sites, shown by the black markers, each with three sectors (surrounding green cells), are considered. Frequency reuse of one is assumed, as is common in contemporary 3/4G network deployments. The inter-site distance (ISD) typically depends on the propagation environment e.g. (dense) urban, suburban, rural, and the spatial traffic intensity. The assumption in this study is that the NB-IoT network is deployed on existing sites (for GSM/UMTS/LTE). Hence, the typical ISDs proposed by 3GPP in [54][55] for specific propagation environments are considered as shown in Table 4-1.

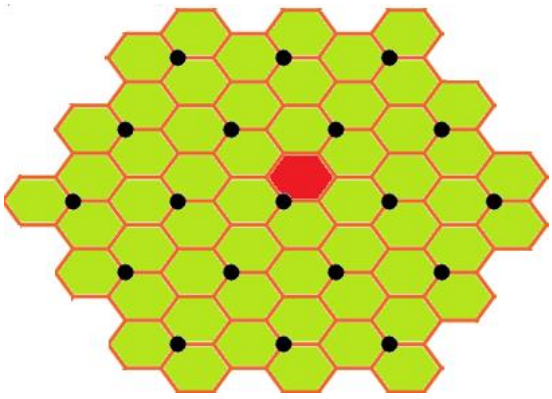

Figure 4-4. Hexagonal grid of radio cells. Cell sites are represented by the black markers. The simulation is executed for only UEs in the radio coverage of the red cell. The green cells represent the interfering cells.

Table 4-1. ISD assumptions for different propagation environments (cell range ≈ 0.67 x ISD).

Propagation environment	ISD (km)	Cell range (km)
Dense urban	0.5	0.34
Urban	1.732	1.16
Suburban	3.2	2.14
Rural	7.5	5

As stated in Section 4.3.1, for simplicity, only a single-cell simulation is executed i.e. for the UEs in the radio coverage of the red cell in Figure 4-4. However, cell selection (see Section 4.3.3) and interference calculation (see Section 4.3.5) is done considering all the 57 cells, which includes 56 interfering cells.

4.3.2.2 Energy distribution network layout

There are three main components modelled in the energy distribution network – the households (each with a smart meter installed), the substations and the distribution feeders connecting multiple households to a substation. The general distribution network topology shown in Figure 4-5 [56] in The Netherlands is the basis for the model developed in this section. The distribution network consists of a primary ring of substations which convert medium voltage (MV, 10 kV) to low voltage (LV, 0.4 kV or 0.2 kV) for households. From each of the substations, feeder cables originate in a radial topology with a certain number of household connections per feeder. Note that a single connection point on a feeder may serve more than one household. Analogous to radio network planning, distribution network planning mainly involves deciding the number of substations (analogous to cell sites) to be installed in an area, the number of feeders per substation, the feeder length (analogous to the cell range) and the number of connections per feeder. The decision is based on several factors such as cable losses and voltage violation limits [56]. Table 4-2 [56][58] lists the average values of the key distribution network parameters for different environments. Note that an 'environment' is characterized by the following distinctive features:

1. Propagation aspects, such as the height of buildings and physical topography, which influence the radio network ISDs shown in Table 4-1.
2. Average household density, i.e. the number of households per km^2, which influences the number of smart meters in an area e.g. that covered by a radio cell.

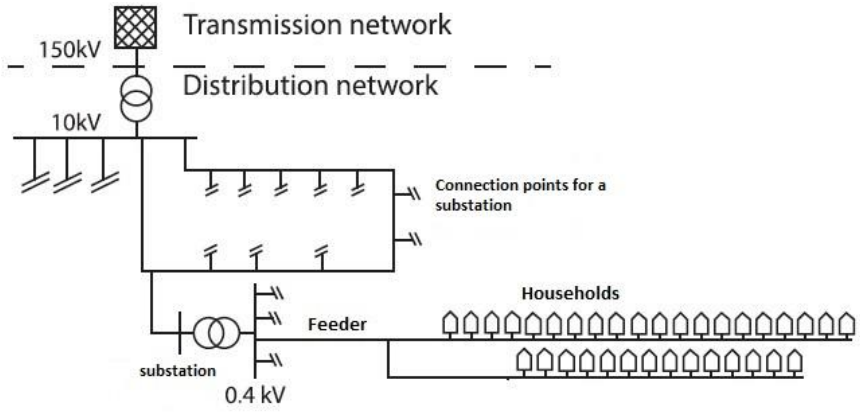

Figure 4-5. General energy distribution network topology in The Netherlands [56]. The different substations are connected in a ring. Feeders connect different households in a radial topology.

Table 4-2. Typical energy distribution network parameters for different environments in The Netherlands. (HH refers to 'household').

Environment	Average HH density (per km^2)	Feeder length (km)	Connections per feeder	HHs per connection	HHs per feeder	Feeders per substation	HHs per substation
Dense urban	2272	0.35	31	2.5	77	9	693
Urban	1500	0.4	30	2	60	8	480
Suburban	350	0.6	22	1.5	33	5	165
Rural	50	0.8	8	1	8	3	24

Similar to a radio cell coverage area, practical substation coverage areas are also irregular because of restrictions/limitations in laying of feeder cables. Thus, feeder cables connecting households may not necessarily follow a straight line from the substation. However, given the radial topology, for modelling purposes, we can represent the maximum coverage area of a substation for a given feeder length, using a circle (approximated as a hexagon) with the substation at the center and the radius equal to the feeder length. Any irregular coverage area will naturally fall within this circular area, for the same feeder length.

Based on the above assumptions, a hexagonal layout can be used for the energy distribution network, where each hexagon represents the coverage area of a substation. In order to determine the effective feeder length (hexagon radius) for such a substation, we first determine the value of the hexagonal area, termed as the effective substation coverage area, such that the number of households covered in the area equal the expected number as per the last column in Table 4-2. This is essential to satisfy the capacity constraints of the substation. Using this area, the hexagon radius, termed as the effective feeder length, is determined. The effective feeder lengths derived in this manner, for all the environments, is shown in Table 4-3.

Table 4-3. Derivation of effective feeder length of a substation for the different environments.

Environment	Average HH density (per km^2)	Effective substation coverage area (km^2)	Effective feeder length (km)
Dense urban	2272	0.31	0.35
Urban	1500	0.32	0.35
Suburban	350	0.47	0.43
Rural	50	0.48	0.43

Combining the relevant parameters, i.e. the ISD and effective feeder length, for specific environments from Table 4-1 and Table 4-3, the radio and energy distribution network layout is generated and illustrated in Figure 4-6. For each environment, the radio and energy distribution network, both represented by a hexagonal grid, are overlaid over each other. Depending on the environment, the number of substations covered within a radio cell increases in the order: dense urban, urban, suburban and rural. This is a net effect of two effects: the ISD of the radio cell, and the household density, both of which depend on the environment, but to a different degree. The network load experienced by the radio cell depends on the product of the number of substations covered in the radio cell and the number of households per substation. It can be noted from Table 4-2 that the latter decreases when moving from dense urban to rural environment.

The figure also shows that the degree of overlap between a substation and radio cell is not uniform across all radio cells, particularly in the dense urban and urban environments. This, in turn, may impact the network load in a radio cell during a power outage scenario which affects one or more substations selected randomly. In order to model this variability in the network load in a single radio cell simulation, the relative position of one of the grids with respect to the other is chosen randomly in certain simulation snapshots, as part of Step 1 in the simulation flow described in Section 4.3.1. Consequently, this would result in varying overlap between the radio cells (including the one selected for simulation) and the substation cells.

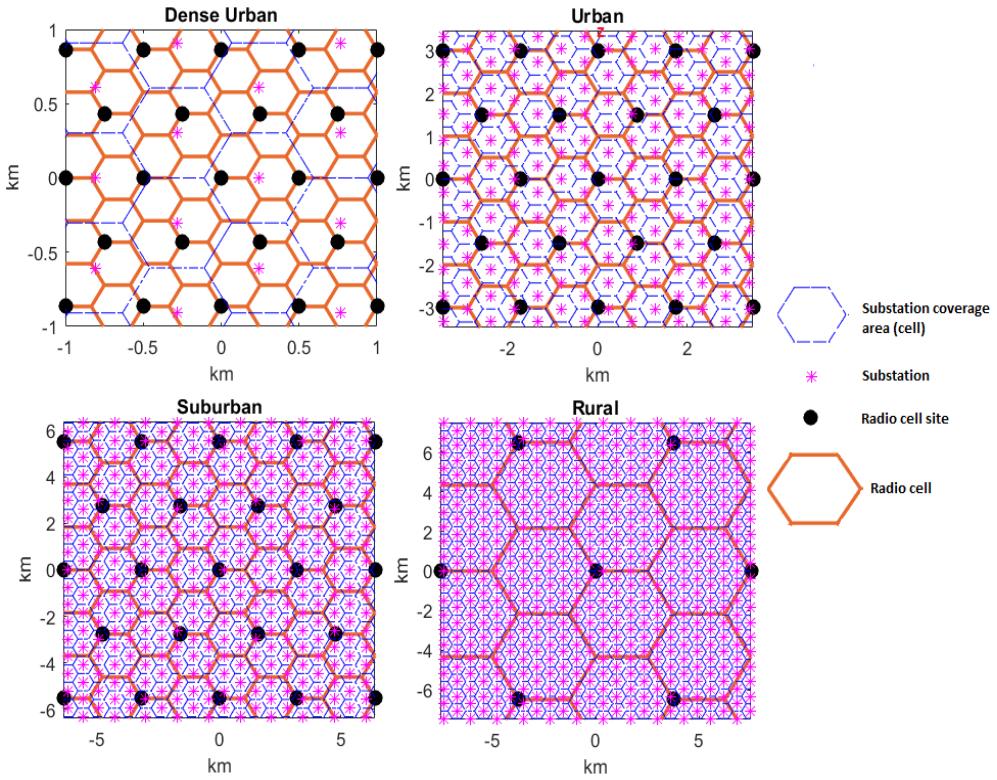

Figure 4-6. Radio network and energy distribution network layout for different environments.

4.3.3 Wireless channel propagation model

The wireless channel propagation model parameters and equations, based on assumptions of 3GPP [5][54] and in [59], are summarised in Table 4-4.

Table 4-4. Summary of the propagation parameters for different environments.

Environment (Frequency: 800 MHz, 3-sector macro cell)	**Path loss equation**	**Shadowing[8] standard deviation (dB)**	**Building penetration loss [8] (dB)**	**Base station antenna specifications**		**Multipath channel model**
				Pattern	**Height (m)**	
Dense urban (DU)	119.8 + 37.6* log10(D) D(km)	10	Based on adapted COST 231 NLOS model (see B)	3D model proposed in [60] (Maximum gain: 18 dBi)	30	Typical Urban (TU), 20 taps, 0 Hz Doppler Spread
Urban (U)						
Suburban (SU)	103.8 + 33.6* log10(D) D(km)	8			45	
Rural (RU)	94.6 + 34.1*log10(D) D(km)	6				Rural area channel model, 6 taps, 0 Hz Doppler spread

The large-scale signal fading is characterised by an outdoor path loss, outdoor shadowing, building penetration loss and the antenna gain. The base station antenna height, specified in the table, is used in calculating the antenna vertical tilt, such that the antenna main lobe points towards half the cell range. The vertical tilt is the used to determine the antenna gain using the three-sector antenna model proposed in [60]. Note that, for the UE, an omnidirectional antenna pattern (0 dBi gain) is assumed with an antenna height of

[8] Inter-site correlation of 0.5 and intra-site correlation of 1 is assumed.

1.5 m. Small-scale fading is determined using multipath channel models. Since smart meters are static devices and will only experience very slow channel variations, it is reasonable to assume a 0 Hz Doppler spread in the multipath channel model. Every UE is initialised with a small-scale fading gain, picked from a distribution characterised by the channel model, independently in the downlink and uplink, in each simulation snapshot.

Applying the channel propagation models given in the above table, the coupling loss between the UE and eNodeB at each cell in the downlink can be derived as follows:

$$\mathrm{CL_{DL}}\ (\mathrm{dB}) = \mathrm{PL}\ (\mathrm{dB}) + \mathrm{X}\ (\mathrm{dB}) + \mathrm{BPL}\ (\mathrm{dB}) - \mathrm{G_{MP,DL}}(\mathrm{dB}) - \mathrm{G_{Ant,eNB}}(\mathrm{dB}) - \mathrm{G_{div}}(\mathrm{dB}) \qquad (4.1)$$

with

- **CL_{DL}** the coupling loss in the downlink
- **PL** the path loss
- **X** the shadowing loss
- **BPL** the building penetration loss
- **$G_{MP, DL}$** the multipath gain in the downlink
- **$G_{Ant, eNB}$** the antenna gain at the eNodeB
- **G_{div}** the diversity gain

To derive the uplink coupling loss, **$G_{MP, DL}$** is replaced with **$G_{MP, UL}$** in the above equation. Note that in the equation, a diversity gain (**G_{div}**) of 3 dB is assumed based on a configuration of two antennas at the eNodeB and one antenna at UE. This is based on the assumption that there are two independent channels available as derived from the product of the number of transmit and receive antennas. Consequently, a two-fold (i.e. 3 dB) diversity gain in the signal power is obtained at the receiver.

The UE selects the serving cell, amongst the 57 candidate cells, as the one with the minimum **CL_{DL}**. In a practical scenario, the UE selects the cell with the highest RSRP value. Assuming that the transmit power settings for the reference signal are same for all the cells, this is the same as selecting the cell with the minimum **CL_{DL}**, as is modelled in this study.

An example coverage map for the central site resulting from the cell selection process described above is shown in Figure 4-7 for an urban environment.

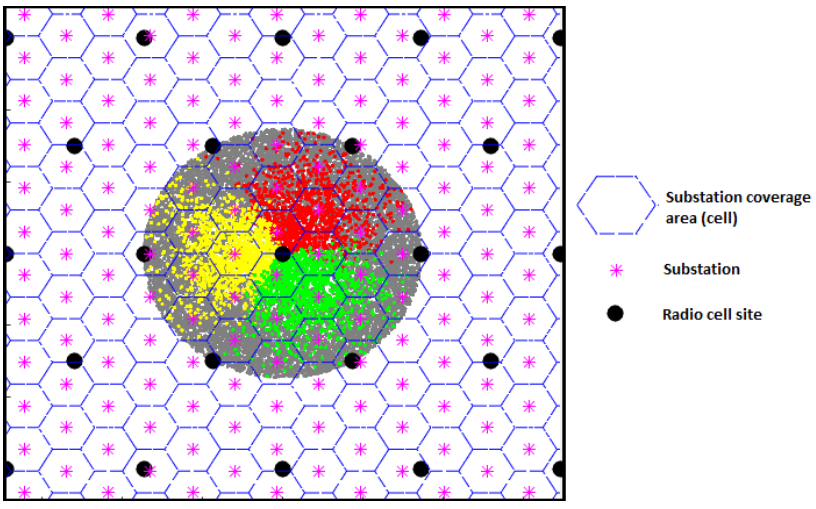

Figure 4-7. Radio and energy distribution network coverage map for a region around the central site, shown for an urban environment.

The red, green and yellow coloured pixels represent the UEs which selected the respective three cells of the central site as their serving cell. The remaining grey pixel-UEs are served by one of the remaining 54 cells. It can be noted that the actual radio cell coverage area is irregular compared to the hexagonal representation due to random shadowing, multipath and building penetration loss effects. Consequently, a subset of UEs

covered by a radio cell also exists outside the simplistic hexagonal cell boundary indicated in Figure 4-4. Therefore, for a more accurate depiction of a radio cell's coverage area, UEs need to be dropped in a much larger area than that of a hexagon, as depicted in the figure.

4.3.4 Traffic model

Although the focus of this study is on the performance of power outage notifications from smart meters, it is necessary to include, in the traffic model, not just the power outage notification traffic but also the regular traffic, e.g. meter-readings, from the smart meters. This is because a power outage can, in principle, occur when a smart meter or a set of smart meters are attempting to send regular traffic over the network, thus possibly impacting the performance of the power outage notification traffic. To represent the regular traffic generated from the smart meters, we consider only the uplink meter-reading reports since, from the traffic characteristics shown in Chapter 2, it is most significant in terms of the average network load. Table 4-5 lists the characteristics of the two traffic types simulated.

Power outage notification is the high-priority traffic which may be triggered at any instance (uniformly distributed) within the hour. From the time of event (power outage/restoration) occurrence, there is a time window of 10 seconds during which power outage notifications from all the affected smart meters are generated. This is because, typically, power outages propagate through the grid network which results in a time lag, represented by the time window, until the power outage is detected by the smart meter at all the affected households.

The meter-reading reports are assumed to have a periodicity and an end-to-end transfer delay requirement of one hour. To avoid high instantaneous network load, the traffic from all the smart meters is uniformly spread within a one-hour window. A similar assumption is made in [64], though with a much lower periodicity of one minute in order to determine the worst-case network capacity requirement. Since meter readings are not time-critical, it is dropped if a power outage notification is to be sent at the same time. The network scheduler, described in Chapter 5, also prioritises the meters attempting to send emergency data over those sending regular meter-reading data. Note that the device can indicate the type of data to the network at the start of the random access process.

Table 4-5.Traffic model assumptions.

Traffic type	Payload size (bytes)	Arrival distribution (from all smart meters)	Comments
Power outage notification	25	Beta-distributed (α = 3, β= 4) within [0,10] seconds [64], measured from the time of event occurrence	Power restoration event, with a similar traffic model, is not separately simulated since the time between power outage and restoration is assumed to sufficient enough that its performance is statistically similar.
Meter-reading reports	200	Uniformly distributed within a time window of one hour	The payload size assumes four sub-interval readings for every hour [9]. At the turn of every hour, each meter generates the reading of the previous hour and transmits this reading at a randomly selected time in the upcoming hour.

The traffic model in Table 4-5 is only applied to UEs at certain coordinates of the network layout, explained as follows:

1. A set of UEs is first dropped uniformly around the central site in a circle of radius equal to the ISD of the eNodeBs, as shown in Figure 4-7. A larger radius was not selected since the additional UEs dropped beyond the ISD would be closer to the network layout boundary than to the central site. This would result in less accurate cell selection for those UEs since their 'best' server cells may actually be those beyond the considered network layout boundary.
2. Since only a single cell is simulated, the subset of UEs covered by the red cell in Figure 4-7 is then finally selected for further simulation.

3. The above red pixel-UEs are grouped as per the substations (cells) they are covered in.
4. The traffic model of the meter-reading reports is applied to the red pixel UEs.
5. During a power outage scenario affecting one or more substations, the power outage notification traffic model is then additionally applied to the UEs covered by the affected substation (s).

4.3.5 PHY layer model

The PHY layer model is responsible for abstracting the link performance based on the input SINR and the chosen MCS. The input SINR is inclusive of the gain achieved due to added repetitions. Under ideal channel estimation at the receiver, in an AWGN channel, the SINR increases by the same factor as the number of repetitions. For example, two repetitions result in doubling of the SINR i.e. it produces a 3 dB gain.
The SINR computation corresponding to the simulation cell eNodeB and a UE **k** served by the simulation cell, for downlink and uplink is as follows:

$$SINR_{DL,k} = \frac{P_{tx,eNB} \times CL_{DL,k} \times N_{Rep,DL,k}}{I_{UE,k} + NF_{UE} \times N_0 \times 180000} \quad (4.2)$$

$$SINR_{UL,k} = \frac{P_{tx,UE} \times CL_{UL,k} \times N_{Rep,UL,k}}{(i_{eNB} + NF_{eNB} \times N_0) \times BW_{UL,k}} \quad (4.3)$$

with

- $P_{tx,eNB}$ the transmit power (mW) of the eNodeB across the NB-IoT carrier (180 kHz bandwidth)
- $P_{tx,UE}$ the transmit power (mW) of UE **k**
- $CL_{DL,k}$ and $CL_{UL,k}$ the coupling loss between UE **k** and the eNodeB in the downlink and uplink respectively
- $N_{Rep,DL,k}$ and $N_{Rep,UL,k}$ the number of repetitions assigned to UE **k** in the downlink and uplink respectively
- $I_{UE,k}$ the average (over time) downlink interference power (mW) in the 180 kHz bandwidth from the interfering cells at UE **k**
- i_{eNB} the average (over time) uplink interference power density (mW/Hz) from the interfering cells at the eNodeB
- NF_{eNB} and NF_{UE} the receiver noise factor (equivalent to noise figure in dB) at the eNB and UE **k** respectively
- N_o the thermal noise power density (mW/Hz)
- $BW_{UL,k}$ the allocated bandwidth (Hz) in the uplink for UE **k** respectively

Since traffic in the interfering cells is not explicitly simulated, a fixed value of I_{eNB} at the eNodeB and $I_{UE,k}$ at UE **k** is assumed. This fixed value is chosen to be the respective average (over time) interference power values from the interfering cells. For determining I_{eNB} and $I_{UE,k}$, without simulating traffic from the interfering cells, the following assumptions are made:

1. The traffic model in the interfering cells is symmetric to that of the simulation cell.
2. Based on 1. above, the average (over time) uplink interference power at the simulation cell eNodeB is equal to the average (over time) total uplink interference power from the simulation cell to the interfering cell eNodeBs.

A detailed description of the interference calculations and related assumptions are provided in Appendix C.

The general simulation settings that are used as an input in the SINR calculations are listed in Table 4-6. Note that in the in-band and guard-band cases, the total available eNodeB transmit power is 46 dBm which is typically uniformly distributed across the LTE system bandwidth of 50 PRBs, which results in a transmit power per PRB equal to roughly 29 dBm. To provide better downlink coverage for NB-IoT, the transmit power on the NB-IoT carrier (equal to a PRB) is additionally boosted by 6 dB, by sacrificing some power used for the LTE resource elements. Note that since the boosted power comes from 49 PRBs available for LTE, the reduction

in power per LTE resource element will not be significant. With power boosting, the effective transmit power of the NB-IoT carrier is 29 + 6 i.e. 35 dBm. For the stand-alone case, the transmit power assumption is based on a typical GSM deployment scenario on a 200 kHz spectrum [55].

Table 4-6. General simulation assumptions.

Parameter	Value
LTE system bandwidth (for in-band and guard-band mode) GSM bandwidth (for stand-alone mode)	10 MHz (50 PRBs) 200 kHz
eNodeB transmit power for NB-IoT (dBm)	35 dBm (in-band and guard-band) 43 dBm (stand-alone)
UE transmit power (dBm)	23 dBm
Thermal noise power density	-174 dBm/Hz
eNodeB receiver noise figure	3 dB
UE receiver noise figure	5 dB

In system-level simulations, the link performance for any physical channel is typically modelled using a set of SINR versus BLER curves for different MCS indices. The MCS-specific reference SINR-BLER curves used in this thesis for the relevant NB-IoT physical data channels (NPUSCH and NPDSCH) are based on [61] and [62]. These curves are then used to derive the highest MCS that can be used in a physical layer transmission, for a given input SINR, such that the BLER is below the target value of 10 %. This is shown using a SINR-MCS step function in Figure 4-8, as an example for NPUSCH single-tone transmission. Using this step function, the appropriate MCS index for any input SINR can be selected. For example, if the link SINR is 6 dB, MCS index 5 is selected.

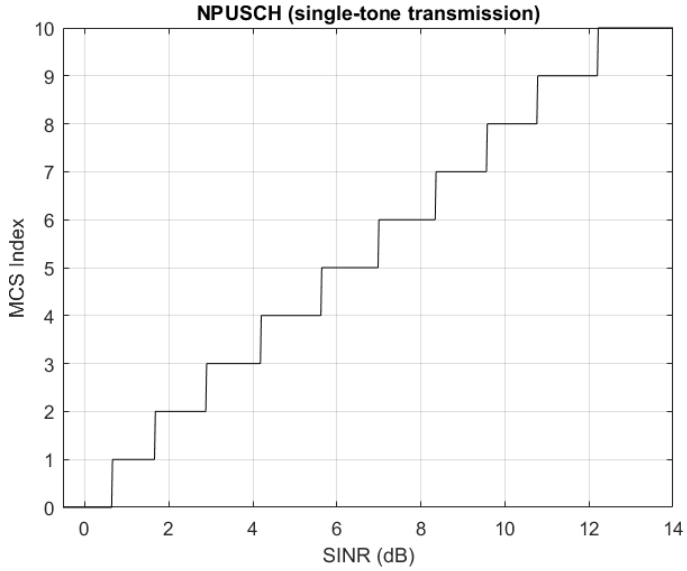

Figure 4-8. SINR-MCS step function (for NPUSCH single-tone transmission).

For a detailed description of how the SINR-MCS step function is derived using the reference SINR-BLER curves, the reader may refer to Appendix D . SINR-MCS curves for the NPUSCH multi-tone transmission and the NPDSCH (as per the deployment mode) are additionally provided in Appendix E.

4.3.6 Link adaptation

The link adaptation model is only applied in the downlink to update the choice of the number of NPDCCH repetitions, NPDSCH MCS and the number of NPDSCH repetitions for a UE. The reason for this is explained below.

As stated in Section 3.2.2, CQI reporting from the UE is not supported in NB-IoT. CQI reporting, in general, is a mechanism by which the UE reports to the eNodeB, the highest MCS suitable for downlink transmissions with a target BLER, e.g. 10 %, based on its SINR measurement. Effectively, this provides information to the scheduler about the downlink channel conditions at the UE which help in the assignment of the appropriate MCS. Since such information is not directly available to an NB-IoT scheduler, it must rely on the ACK/NACK feedback which a UE provides upon successful/unsuccessful decoding of a transport block. Link adaptation plays a role here in providing the necessary logic for updating (or adapting) the scheduler's choice of MCS and the number of repetitions. Link adaptation in the uplink is not considered in this study because of the assumption that the channel is static with a fixed interference level and that UEs always transmit with maximum power. The latter assumption means that the scheduler can determine the uplink coupling loss for a UE using the received power of the preamble, during the preamble transmission phase. The expected uplink SINR is then calculated for the assigned number of subcarriers. Consequently, the appropriate MCS and number of repetitions are assigned. The link adaptation model used in this study is based on the model proposed in [26] for NB-IoT, which has been shown to be more effective than certain alternative approaches. The model is illustrated as a flow diagram in Figure 4-9.

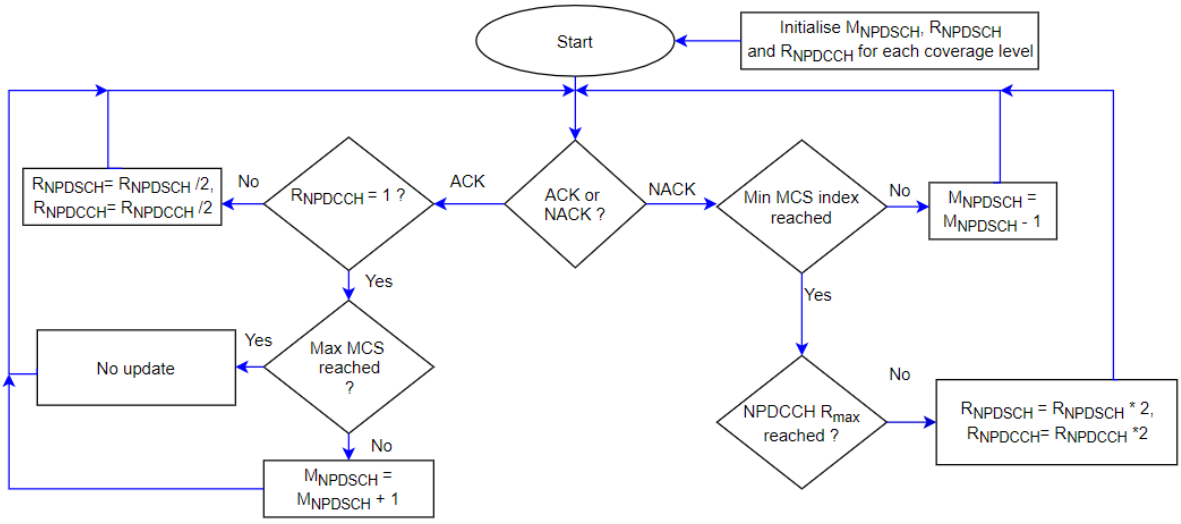

Figure 4-9. Link adaptation flow diagram.

At the start, the NPDCCH MCS index (**M_{NPDSCH}**), the number of NPDSCH repetitions (**R_{NPDSCH}**) and the number of NPDCCH repetitions (**R_{NPDCCH}**) are initialised for each coverage level. A simple scheme for initialising the values is followed. The lowest and highest MCS indices are chosen for the highest and lowest coverage levels respectively. The middle coverage level is initialised with the centremost MCS index. **R_{NPDSCH}** and **R_{NPDCCH}** repetitions are initialised to one, except for the highest coverage level where they are initialised to two. A low value is chosen for faster MCS adaptation. Although the above form of initialisation does not necessarily provide optimal adaptation for all UEs, it has been verified through simulation results that the impact of downlink decoding errors, arising from such an initialisation, on the overall performance is not significant.

Note that the model in Figure 4-9 is applied only for MSG 4 scheduling since prior to MSG 4, the UE would, by default, update its coverage level, which would have a more robust MCS scheme configured if it experiences downlink decoding errors. After MSG 4, link adaptation is not applied since no NPDSCH messages are involved.

For UEs which are to be scheduled for MSG 4, the MCS index and repetition values are initialised as per their updated coverage levels and the flow diagram is followed. This is explained as follows:

1. With every ACK, if the current number of repetitions is greater than one, it is halved. Note that, as stated in Chapter 3, repetitions greater than one are always a multiple of two.
2. If the number of repetitions is already at the minimum value, the MCS index is incremented by one instead until the maximum MCS index is reached.
3. With every NACK, the MCS index is decremented by one until it reaches the minimum value.
4. With further NACKs, the number of repetitions is increased until the data is successfully decoded or the configured maximum number of repetitions i.e. $\mathbf{R_{max}}$, as per the NPDCCH configuration for the coverage level, is reached. In the latter case, a transfer failure is declared.

In general, a fast adaptation approach is followed i.e. an update is made with every ACK/NACK since only a single downlink message transfer is involved.

Using the MCS index derived from the link adaptation model and the knowledge of the size of the message (RAR or MSG 4)[9] in bits, the allocated number of subframes is given in Table 4-7 [47].

Table 4-7. Mapping between MCS index, number of subframes and the TBS in bits for downlink transmission.

MCS	Number of subframes							
	1	2	3	4	5	6	8	10
0	16	32	56	88	120	152	208	256
1	24	56	88	144	176	208	256	344
2	32	72	144	176	208	256	328	424
3	40	104	176	208	256	328	440	568
4	56	120	208	256	328	408	552	680
5	72	144	224	328	424	504	680	
6	88	176	256	392	504	600		
7	104	224	328	472	584	680		
8	120	256	392	536	680			
9	136	296	456	616				
10	144	328	504	680				
11	176	376	584					
12	208	440	680					

4.4 Summary

In this chapter, an overview of the simulation model was presented and each of the interacting components of the model was developed and described. The following model components were presented:

1. The *simulation flow*, which models, via a set of discrete events, the random access process for small data transmission.
2. The *network layout*, which specifies the environment-specific layout of the radio and energy distribution networks.
3. The *wireless channel propagation model*, which specifies, based on the environment input, the coupling loss between the UE and eNodeB.
4. The *traffic model*, which characterises the spatio-temporal traffic arrival process of the power outage notifications and the regular smart meter meter-reading traffic.
5. The *PHY layer model*, which computes the SINR and abstracts the link performance using the BLER metric.
6. The *MAC layer model*, which models link adaptation and scheduling.

Overall, the simulation is carried out as a series of 'snapshots', for a given combination of simulation inputs. Each snapshot is characterised by a random initialisation of the network layout and/or the traffic model. Several snapshots are simulated and relevant KPIs are then statistically derived based on the KPI output of each snapshot.

[9] The MAC PDU of the RAR and MSG 4 is approximately four bytes (32 bits) [50] and 30 bytes (240 bits) [49] respectively.

Chapter 5. Scheduling

In this chapter, certain candidate NB-IoT schedulers are proposed which will be evaluated as part of this study. The aim is to maximise the reliability performance of the power outage notifications, which can be done by maximising the success rate and minimising the 95th transfer delay percentile of successfully received notifications. Before discussing the details of the different scheduling flavours, we will first look at, in Section 5.1 , where and how, scheduling plays a role in influencing the performance of the system with respect to the success rate and the transfer delay. In Section 5.2, we will look at the general steps involved in the scheduling process. The discussion of the different scheduling flavours is then provided in Sections 5.3 and 5.4.

5.1 System description

Figure 5-1 shows a simplified system-level traffic flow within the random access process and the uplink data transmission that follows a successful access. Note that a successful access is declared after receiving MSG 4. There are three phases depicted: preamble transmission, RAR/MSG4 transmission and UL data transmission. For each phase, the relevant physical resources involved are also listed. For clarity, only the main traffic contributing factors have been included i.e. the traffic arising from uplink or downlink decoding errors in any of the phases is not shown.

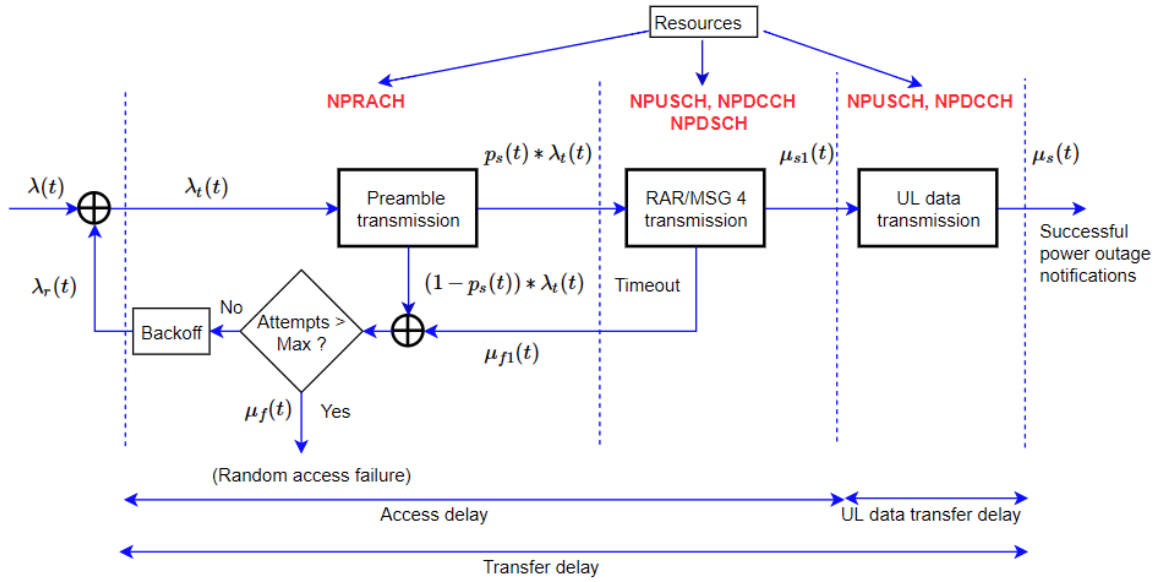

Figure 5-1. Illustration of a simplified traffic flow within the random access process and the uplink data transmission that follows, from a system perspective.

Figure 5-1 is explained as follows:

1. The arrival rate, i.e. the number of arrivals per second, of new traffic is represented by **λ (t)**. Note that **λ (t)** will vary over time because of the nature of the beta-distributed traffic arrivals. Thus, the figure can be considered a snapshot of the system at time **t** and therefore **λ (t)** represents the instantaneous value as is also applicable to the other variables in the figure.
2. The net arrival rate of preamble attempts **λ_t (t)** is the sum of **λ (t)** and **λ_r (t)** which is the rate of preamble reattempts.
3. The fraction of the total number of preamble attempts that proceeds to the RAR/MSG 4 transmission phase, is the preamble detection success probability **p_s (t)**, which primarily depends on the probability of non-collision of the preamble. The collision probability, in general, similar to

Equation (3.1), depends on λ_t **(t)** and the available NPRACH resources expressed in terms of the number of RAOs per second.

4. p_s **(t)** * λ_t **(t)** is the net arrival rate of the traffic which needs to be scheduled for RAR, followed by MSG 4. The net outgoing traffic from the second phase is represented by a serving rate μ_{s1} **(t)** and a timeout failure rate μ_{f1} **(t)**. The serving rate is the number of UEs, per second, which have successfully received MSG 4 whereas the timeout failure rate is the number of UEs timed out per second. μ_{s1} **(t)** and μ_{f1} **(t)** are impacted mainly by the scheduling scheme and the availability of NPUSCH, NPDCCH and NPDSCH resources. For example, with more NPUSCH resources and an effective scheduling scheme that maximises the number of UEs served in a time period, μ_{s1} **(t)** can be increased. For a given p_s **(t)** * λ_t **(t)** and set of window sizes for RAR and MSG 4, an increase in μ_{s1} **(t)** obviously leads to a decrease in μ_{f1} **(t)** which leads to a decrease in λ_t **(t)**.
5. A fraction of the sum of **(1** - p_s **(t))** * λ_t **(t)** and μ_{f1} **(t)** may exhaust its maximum available attempts and is then dropped from the system resulting in an overall failure rate of μ_f **(t)**. The remaining fraction attempts random access after a certain backoff time, with an effective rate λ_r **(t)**.
6. Similar to the RAR/MSG 4 transmission phase, the serving rate μ_s **(t)** in the UL data transmission phase is dependent on the scheduling scheme and the availability of physical resources (NPUSCH and NPDCCH). However, due to the differences in waiting time for scheduling (including for transmission) and delays in the transmission of data, μ_s **(t)** may or may not equal μ_{s1} **(t)**. The mathematical integration of μ_s **(t)** over time ultimately determines how many UEs are successful in the end i.e. the success rate.
7. Depending on the instantaneous failure probabilities in the preamble detection and RAR/MSG 4 scheduling phases, a given UE may end up making a certain number of attempts to be successful in the random access, i.e. the first two phases. The higher the number of random access attempts made, the higher is the magnitude of the access delay.
8. The transfer delay is the sum of the access delay and the UL data transfer delay. Thus, an increase in the access delay due to an increase in the number of random access attempts made can possibly increase the transfer delay as well.

5.2 Overview of the scheduling process

From the random access procedure call flow shown in Figure 3-13 and the system description in the previous section, it can be concluded that there are three scheduling decisions to be made, in the last two phases of Figure 5-1, during the overall data transmission process. They are listed as follows:

1. Scheduling of the RAR in the downlink and MSG 3 in the uplink
2. Scheduling of MSG 4 in the downlink and ACK/NACK in the uplink
3. Scheduling of uplink data containing the power outage notification

Thus, at a given time, there may be a contention amongst the UEs to be scheduled for different types of messages above. Each message type also has an associated delay budget in the form of either specific time windows (order of ms), applicable to the RAR and MSG 4, or an overall end-to-end delay target, e.g. 20 seconds for the power outage notification, applicable to the uplink data. Violations of the RAR and MSG 4 time windows are particularly critical since they would lead to random access reattempts. This could lead to possible transfer failures once the maximum allowed number of random access attempts is exceeded, thereby impacting the success rate. Further, as seen in the previous section, an increase in the number of random access attempts increases the access delay and possibly also the transfer delay.

Based on the above context, we now look at the general steps involved in the scheduling process illustrated in Figure 5-2. The first step in the scheduling process involves *determining the priority order* in which to serve the UEs in the scheduling queue. Each of these UEs would have experienced a certain waiting time until the current scheduling instance and there may be groups of UEs with different delay budgets if there is more

than one message type to be scheduled. Therefore, the prioritisation metric computation, as will be shown in the next section, considers both the UE's waiting time and its respective delay budget.

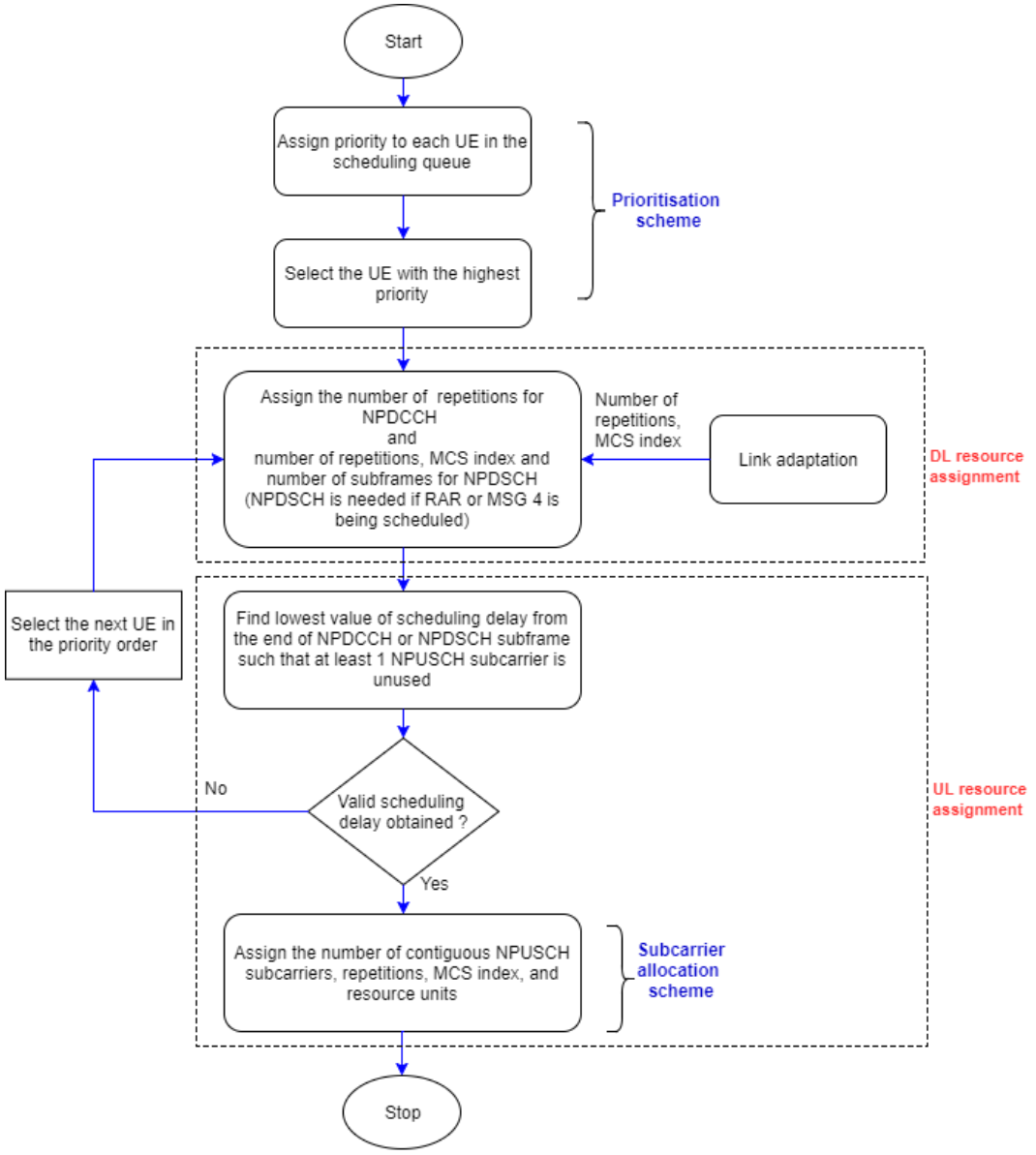

Figure 5-2. General steps in the scheduling process.

Once the priority order is determined, the highest priority UE is selected for scheduling. For the selected UE, the next step is to assign the *downlink* physical layer resources for NPDCCH and NPDSCH (if RAR or MSG 4 is to be scheduled). This includes the required number of NPDCCH repetitions, the NPDSCH MCS index, the number of NPDSCH subframes and repetitions. These parameters are derived from the link adaptation loop as explained in Section 4.3.6.

Next, the *uplink* resources are assigned, which may be for MSG 3, ACK/NACK (for MSG 4) or the uplink data. Firstly, the number of unused NPUSCH subcarriers available to be scheduled and the possible sets of contiguous allocations for single and/or multi-tone transmissions are determined. If we recall the generic uplink scheduling illustration in Figure 3-12, we saw that the *scheduling delay* parameter ($\mathbf{k_o}$) can be used to schedule the uplink transmission to a time instant where one or more subcarriers are unoccupied. Out of the several possible $\mathbf{k_o}$ values, we pick the lowest $\mathbf{k_o}$ such that at the corresponding time instant, at least one subcarrier is available to be scheduled. If no such $\mathbf{k_o}$ exists, then the current UE is dropped and the next one in the priority order is selected. Based on the set of possible contiguous allocations, i.e. one, three, six or

twelve subcarriers, the *subcarrier allocation scheme* determines how many subcarriers are to be allocated to the UE.

Once the number of assigned subcarriers is known, the uplink SINR is calculated. The uplink SINR is dependent on the number of allocated subcarriers since the available UE transmit power is distributed over the allocated subcarriers only. Using the SINR-to-MCS mapping, as shown in Figure 4-8, the appropriate MCS index is chosen. If the SINR is lower than the minimum SINR for the lowest MCS i.e. MCS index 0, then the number of repetitions and the MCS are selected as follows:

1. To minimise the transmission time, only the minimum required number of repetitions is assigned such that the accordingly estimated effective SINR is greater than or equal to at least the minimum required SINR for MCS 0.
2. If the effective SINR calculated above is such that it is also greater than or equal to the minimum required SINR for one or more higher MCS indices, then the highest such MCS index is assigned, otherwise, MCS 0 is assigned.

With the MCS index known, the required number of resource units is determined based on the amount of data (in bits) that is to be transmitted. The mapping between the MCS index, the TBS and the resource units from [47] is used and shown in Table 5-1.

Table 5-1. Mapping between the MCS index, the resource units and the TBS in bits for uplink transmission. For single-tone transmission, MCS indices 11 and 12 are not used [47].

MCS	Number of resource units							
	1	2	3	4	5	6	8	10
0	16	32	56	88	120	152	208	256
1	24	56	88	144	176	208	256	344
2	32	72	144	176	208	256	328	424
3	40	104	176	208	256	328	440	568
4	56	120	208	256	328	408	552	680
5	72	144	224	328	424	504	680	872
6	88	176	256	392	504	600	808	1000
7	104	224	328	472	584	712	1000	
8	120	256	392	536	680	808		
9	136	296	456	616	776	936		
10	144	328	504	680	872	1000		
11	176	376	584	776	1000			
12	208	440	680	1000				

Different flavours of the prioritisation and the subcarrier allocation schemes will be proposed in the following two sections respectively. Each candidate scheduler thus corresponds to a combination of the prioritisation and the subcarrier allocation schemes.

5.3 Prioritisation schemes

In this section, we will discuss possible schemes for UE prioritisation, which covers the first two steps in the scheduling process of Figure 5-2.

When addressing QoS constraints such as the end-to-end delay budgets of packets, the class of schedulers based on the *Earliest Due Date First (EDDF)* prioritisation principle has been shown in existing studies ([30]-[32]) to be effective in achieving lower packet loss rates and 95th transfer delay percentile, compared to other schedulers such as opportunistic scheduling. Note that in these schedulers, the packet loss rate is reflective of the rate of delay budget violations since packets in the queue which exceed their delay budgets are dropped. As described in the previous section, delay budget violations of the RAR and MSG 4 are important since they can directly impact both the success rate and the transfer delay. In view of this, the EDDF principle is considered as a candidate prioritisation criterion.

The main drawback, however, of the EDDF criterion is that it does not consider the expected transmission time, which may vary between UEs depending on the channel conditions and the amount of data to send. The transmission time may impact how long the resources are blocked for other UEs, particularly if the load is high, thereby impacting their waiting times and the likelihood of delay budget violations. In view of this, the well-known *Shortest Processing Time First (SPTF)* principle [65] is considered as a second candidate prioritisation criterion. In a generic sense, the SPTF principle works by scheduling, until completion, first those tasks in the queue that require the least processing time. The motivation is to minimise the average waiting time of the tasks, though at the expense of a higher waiting time for tasks with long processing times. For the current scheduling process, tasks refer to UEs and the processing time refers to the expected transmission time inclusive of both the downlink and the uplink components. Note that the scheduling process is non-pre-emptive, i.e. once a UE is scheduled to transmit a certain amount of data, its transmission (processing) continues until completion and cannot be suspended even if a UE (task) with a shorter processing time arrives in the interim. This is why SPTF is chosen instead of a pre-emptive one such as the Shortest Remaining Time First (SRPTF) principle [66].

The SPTF principle, by definition, does not consider the delay budget of UEs, in contrast to the EDDF principle. This means that, even if a UE is close to its due date, it may still not be assigned high priority if there are other UEs with shorter processing times. Thus, looking at the EDDF and SPTF schemes, it can be remarked that the benefits of one scheme complement the drawbacks of the other. In view of this, a combination of these schemes is considered as the third candidate prioritisation criterion. The formulation of the three candidate prioritisation schemes is discussed as follows.

1. **EDDF**

The first candidate prioritisation scheme is based on the EDDF principle as proposed in [31], which basically assigns a priority to a UE inversely proportional to the time left until its due date and directly proportional to its waiting time. The priority metric for UE **k** is calculated as follows:

$$P_{EDDF,k} = \frac{W_k}{D_k - W_k} \tag{4.4}$$

with

- $\mathbf{W_k}$ the waiting time (ms) of the Head-of-Line (HOL) packet in the buffer of UE **k**. Depending on the state of the UE, this refers to the time elapsed in the RAR/MSG 4 timers or the time elapsed since the last preamble from the UE arrived at the eNodeB[10]
- $\mathbf{D_k}$ the delay budget (ms) of UE **k**. Depending on the state of the UE, this would refer to the configured RAR/MSG 4 time windows or the end-to-end delay budget of the uplink data, e.g. 20 x 10^3 ms for the power outage notification

In the above equation, the $\mathbf{W_k}$ term in the numerator is present to make the priority metric stronger as the UE approaches its due date. At any given scheduling instance, the UE with the highest priority is selected. Note that there may be cases where multiple UEs have the same priority value because of identical values of the waiting time and the delay budget. A typical example is when a set of UEs complete their preamble transmissions at the same time and consequently start the RAR timer. In such scenarios, a random selection amongst the 'tied' UEs is made.

2. **SPTF**

$$P_{SPTF,k} = \frac{1}{T_{DL,k} + T_{UL,k}} \tag{4.5}$$

with

- $\mathbf{T_{DL,k}}$ the expected transmission time (ms) for UE **k** in the downlink
- $\mathbf{T_{UL,k}}$ the expected transmission time (ms) for UE **k** in the uplink

[10] Since the exact time of arrival of the packet and the number of random access attempts made are not conveyed by the UE during the random access process, the scheduler uses the time of arrival of the latest preamble as a reference.

For MSG 3 and uplink data transmission, the calculation of $\mathbf{Tx}_{\mathbf{UL,k}}$ assumes a multi-tone transmission and that all twelve subcarriers are scheduled, i.e. with maximal resource assignment. Note that the actual number/granularity of subcarriers allocated to a UE may be different at a scheduling instance, depending on the number of available subcarriers and the subcarrier allocation scheme.

The general equations for $\mathbf{T}_{\mathbf{DL,k}}$ and $\mathbf{T}_{\mathbf{UL,k}}$ are shown below:

$$T_{UL,k} = R_{NPUSCH,k} * N_{RU,k} * T_{RU,k} \quad (4.6)$$

$$T_{DL,k} = R_{NPDCCH,k} + SD_k + R_{NPDSCH,k} * N_{SF,k} \quad (4.7)$$

with

- $\mathbf{R}_{\mathbf{NPUSCH,k}}$ the number of repetitions of the NPUSCH for UE **k**
- $\mathbf{N}_{\mathbf{RU,k}}$ the number of required NPUSCH resource units (assumed to be one in case of ACK/NACK transmission in the uplink) for UE **k**
- $\mathbf{T}_{\mathbf{RU,k}}$ the time duration of the NPUSCH resource unit for UE **k**, assuming 15 kHz subcarrier spacing
- $\mathbf{R}_{\mathbf{NPDCCH,k}}$ the number of repetitions of the NPDCCH for UE **k**
- $\mathbf{SD}_{\mathbf{k}}$ the overall scheduling delay for UE **k**
- $\mathbf{R}_{\mathbf{NPDSCH,k}}$ the number of repetitions of the NPDSCH for UE **k**
- $\mathbf{N}_{\mathbf{SF,k}}$ the number of required NPDSCH subframes for UE **k**

3. **EDDF with SPTF (EDDF-SPTF)**

In the third candidate prioritisation scheme, Equations (4.4) and (4.5) are effectively combined. The priority metric for UE **k** is calculated as follows:

$$P_{EDDF-SPTF,k} = P_{EDDF,k} \text{ x } P_{SPTF,k} \quad (4.8)$$

i.e.,

$$P_{EDDF-SPTF,k} = \frac{W_k}{(D_k - W_k) * (T_{DL,k} + T_{UL,k})} \quad (4.9)$$

Note that due to the EDDF term, the limitation of the stand-alone SPTF scheme is addressed since this term will dominate as the UEs approach their due dates.

Fundamentally, a trade-off between priority to a UE's due-date and priority to a UE's transmission time is involved in the choice of the three schemes discussed so far. With the EDDF scheme, there is a high priority to a UE's due-date and low priority to its transmission time, whereas the opposite is true with the SPTF scheme. The EDDF-SPTF scheme, on the other hand, aims to achieve a 'common ground' between these extreme cases.

5.4 Subcarrier allocation schemes

In this section, we will discuss possible schemes for subcarrier allocation in the uplink, which is the last step in the scheduling process of Figure 5-2. Three candidate subcarrier allocation schemes are considered for comparison. In each of these schemes, a certain degree of trade-off between the resource assignment granularity/scheduling capacity and the achieved bit rate of UEs, particularly those with high SINRs, is involved. The schemes are explained as follows.

1. **Least Granularity Allocation (LGA)**

The first approach is known as the ''Least Granularity Allocation (LGA)" scheme in which the maximal possible number of subcarriers (out of one, three, six or twelve) from the available subcarriers is allocated. The resources so assigned have the least possible granularity. This approach tries to maximise the bit rate for the selected UE in its immediate scheduled transmission. However, the drawback is that it limits the number of UEs that can simultaneously utilise the uplink resources which could lead to a high waiting time and possible violation of the delay budgets.

It can be noted that for a certain SINR range in the low SINR region of the Shannon capacity curve, the spectral efficiency (bps/Hz) variation with SINR (in dB) is approximately linear. Thus, if the bandwidth is increased by a certain factor resulting in a similar increase in the SINR (since the signal power is now distributed over a lower bandwidth), then the spectral efficiency also increases roughly by the same factor, thereby not affecting the bit rate much. Further, the practical achievable spectral efficiency which is less than the Shannon capacity increases, with SINR, in a step-wise manner since only a discrete set of MCSs are available. Thus, in practice, the gain in spectral efficiency for an increase in SINR may be larger than in the ideal scenario. The "Min-Max Allocation (MMA)" scheme, explained next, exploits this aspect to reduce the waiting time of other UEs.

2. **Min-Max Allocation (MMA)**

The basic idea of the ''Min-Max Allocation (MMA)" scheme is to assign the minimum (Min) number of subcarriers that could provide the maximum (Max) bit rate for the selected UE.

Firstly, all the possible allocation options for the number of subcarriers are considered from the available number of subcarriers. With the SINR calculated from each option, the MCS index, repetitions and resource units are derived. Using these derived parameters, the effective transmission time (same as $\mathbf{T_{x,\,UL}}$ in Equation (4.6)) is calculated and the option with the minimum effective transmission time i.e. the maximum bit rate is selected. In case of a tie, the option with the lower number of subcarriers is chosen. By doing so, the waiting time of the other UEs is reduced compared to that in the LGA scheme.

It can be noted that in both the LGA and MMA schemes, the subcarrier spacing used is 15 kHz. The 3.75 kHz subcarrier spacing has not been additionally considered since there would be a need to allocate a guard band to preserve orthogonality and reduce adjacent channel interference between UEs transmitting with different subcarrier spacing.

An illustration of how the MMA scheme exploits the linear region of the Shannon capacity curve and the discrete nature of MCSs explained earlier, is shown using Figure 5-3. The figure shows the maximum achievable bit rate for the four subcarrier allocation options, expressed as a function of a reference uplink SINR assuming an assignment of twelve subcarriers. In the calculation of the number of required resource units, a data size of 200 bits, which equals the power outage notification payload size, has been assumed.

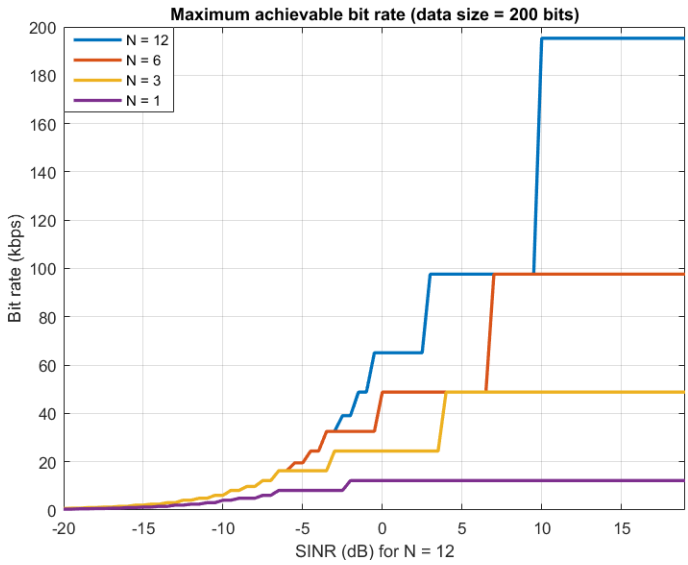

Figure 5-3. Comparison of the maximum achievable bit rate using a data size of 200 bits for different number (N) of assigned subcarriers.

We can see that for low SINRs (< - 5 dB), the bit rate is identical for N = 12, 6 and 3, indicating that the increase in the SINR with lower bandwidth has compensated for a possible reduction in the bit rate, through an increase in the spectral efficiency. In this region, the MMA scheme will tend to choose N = 3. Further, we

see that the bit rate increase occurs in steps, which is mainly due to the availability of only a discrete set of MCSs. At higher SINRs i.e. in the non-linear region, there is a visible gain in the bit rate with an increase in the bandwidth i.e. a higher value of N. In this region, the MMA scheme would tend to choose the highest value of N, similar to the LGA scheme. However, in practice, a high-SINR UE may not achieve the same optimal bit rate during its transmission in the MMA scheme as it does in the LGA scheme. This is because low-SINR UEs may be occupying part of the uplink resources thereby limiting the scheduler from assigning the maximal number of subcarriers to high-SINR UEs.

3. **Maximum Granularity Allocation (MGA)**

The motivation behind this scheme is to maximise the scheduling capacity. This is done by allocating a single subcarrier with a subcarrier spacing of 3.75 kHz, which represents the maximum possible granularity in the uplink. This would allow a maximum of 48 UEs to transmit simultaneously. The trade-off, however, is the non-optimal bit rate for UEs, particularly with high SINRs, and more so when there are not enough UEs with data to transmit, to even utilise all resources.

Considering the above-mentioned possible prioritisation and subcarrier allocation schemes, the following combinations i.e. candidate schedulers, are proposed:

1. EDDF with MGA
2. SPTF with MGA
3. EDDF-SPTF with MGA
4. EDDF-SPTF with MMA
5. EDDF-SPTF with LGA

The motivation behind the above combinations is as follows. The EDDF and SPTF schemes, as highlighted earlier, ignore the possible impact on delay budget violations due to the different transmission time of UEs and the time remaining until the due date respectively. This impact is expected to be more significant if a less granular resource assignment is done in the uplink, such as in the LGA and MMA schemes. Therefore, to minimise this, the MGA scheme is chosen as the suitable combination for EDDF and SPTF prioritisation schemes. Consequently, for LGA and MMA schemes for uplink subcarrier allocation, EDDF-SPTF naturally is the most suitable combination since it aims to address the above drawbacks of both EDDF and SPTF schemes. Additionally, the combination of EDDF-SPTF and MGA is also considered since the transmission time can potentially still impact performance in the MGA scheme due to the low bit rate of UEs. The performance evaluation of these candidate schedulers will be presented in Chapter 6.

5.5 Summary

One of the research objectives of this thesis is the design of suitable NB-IoT schedulers for ORM-based network traffic. To this extent, in this chapter, certain candidate schedulers were proposed, with the aim of maximising the reliability performance. In the design of the scheduler, two key steps in the scheduling process were highlighted, namely, UE prioritisation and the subcarrier allocation in the uplink. A key aspect of the random access process in NB-IoT involves a time-window based scheduling amongst contending UEs, including those UEs waiting to transmit regular data. Additionally, the transmission time of the scheduled UE impacts how long certain resources are blocked for other UEs, thereby affecting their waiting times. Considering these two aspects, candidate *prioritisation* schemes based on the EDDF and SPTF principles respectively, were proposed. However, in view of their inherent 'complementary' benefits and drawbacks, a third scheme based on the combination of the two principles was also proposed. Overall, a trade-off between priority to a UE's due-date and priority to a UE's transmission time is involved in the choice of the three schemes. With regard to the *uplink subcarrier allocation*, three candidate schemes were proposed, which involve a trade-off between the resource assignment granularity and the achieved bit rate for high-SINR UEs, each to a different degree. From these candidate prioritisation and uplink subcarrier allocation schemes, five suitable combinations i.e. candidate schedulers were derived, which will be evaluated in Chapter 6.

Chapter 6. Simulation results and analysis

6.1 Introduction

In this chapter, the performance evaluation of NB-IoT under different network configuration, deployment and power outage scenarios is presented. The main objectives are as follows:

1. To evaluate the impact on network performance of ORM, e.g. in terms of the reliability, due to network deployment in different environments and NB-IoT deployment modes.
2. To evaluate a set of strategies that can help to maximise network performance either by optimising the available physical resources of the carrier or by addition of certain resources.

With respect to 2. above, the focus in this chapter will be on the optimisation of the available resources since it is cost-effective for an operator. The network configuration has different components, such as the NPRACH resource configuration and scheduling schemes, that may need to be optimised to achieve a globally optimal configuration. Alternatively, this can be viewed as a case of multi-dimensional optimisation, where each component represents a dimension in the optimisation space. The basic approach followed is to evaluate the sensitivity of the relevant KPIs and the associated trade-offs, due to unilateral changes in the relevant components. Based on the unilaterally identified optimal settings, a final near-optimal configuration is then derived. Apart from the optimisation of available resources, we will look at a dual-carrier operation to evaluate the performance gain achieved by the addition of resources.

The outline of this chapter is as follows. In Section 6.2, the relevant KPIs used for performance evaluation are defined. In Section 6.3, the simulation settings of a set of baseline scenarios, for evaluation of 1. above, are discussed. In Section 6.4, the results of these baseline scenarios are presented, which then form a basis for the sensitivity analysis presented in Section 6.5. The evaluation of a dual-carrier operation is presented in Section 6.6. Finally, a summary is provided in Section 6.7.

6.2 KPIs

As stated in Chapter 1, the maximisation of reliability is important for the better accuracy of power outage localisation. Thus, reliability is the target KPI to be optimised from a network point-of-view. Since reliability is inherently a function of the success rate and the transfer delays of successfully delivered power outage notifications, it is essential to analyse these two aspects as well. With regard to the latter, we will look at the 95th percentile value of the transfer delays of the successful notifications. The formulations of the relevant KPIs are provided as follows.

6.2.1 Reliability

Reliability, in the context of ORM, is defined as the percentage of generated power outage notifications successfully delivered within the transfer delay target. Here, a transfer delay target of 20 seconds is assumed, as specified in Table 2-2. Thus, the reliability can be expressed as follows:

$$\text{Reliability}\ (\%) = \text{Success rate}\ \text{x}\ \Pr\{\text{D}_{\text{transfer, success}} < 20\ \text{s}\} \tag{6.1}$$

with

- **Success rate** the percentage of generated power outage notifications successfully received at the base station
- $\mathbf{D_{transfer,\ success}}$ the transfer delay (in seconds) of a successfully delivered power outage notification

6.2.2 Success rate

The success rate is computed as follows:

$$\text{Success rate } (\%) = \frac{N_{success} \text{ x } 100}{N_{total}} \quad (6.2)$$

with

- $N_{success}$ the number of power outage notifications successfully received at the base station
- N_{total} the total number of power outage notifications generated from the smart meters

It is assumed that each smart meter affected by a power outage generates a single notification. Thus, N_{total} equals the total number of smart meters affected by the power outage.

6.2.3 95th transfer delay percentile

The 95th transfer delay percentile is measured as the 95th percentile value of the transfer delay, in the radio network, determined over all successfully received power outage notifications. For a successfully delivered power outage notification, the transfer delay is computed as follows:

$$D_{transfer,success} = t_{packet,received} - t_{packet,arrived} \quad (6.3)$$

with

- $t_{packet,\ received}$ the time instance when the packet containing the power outage notification is received at the base station
- $t_{packet,\ arrived}$ the time instance when the packet containing the power outage notification arrives at the UE buffer

6.3 Baseline scenarios: description

To evaluate the network performance in different environments and NB-IoT deployment modes, we first consider a baseline network configuration and a set of scenarios referred to as the 'baseline scenarios'. Each baseline scenario is defined by a combination of an environment and a deployment mode from the following set:

- **Environment**: rural, suburban, urban and dense urban
- **Deployment mode**: in-band, guard-band and stand-alone

Table 6-1 summarises the settings of the baseline network configuration used for the evaluation of the above scenarios. They are described as follows:

1. The first row shows the coupling loss thresholds used for the three coverage levels (see Section 3.5.2 for the definition of a coverage level) configured in the network. The threshold values were chosen such, as to reflect assumptions on the typical distribution of UEs over the different coverage levels shown in [67]. With this distribution, around 90% of the UEs were in coverage level 1 and 7% in coverage level 2, corresponding to the given reference.
2. The carrier operation mode used is 'single-carrier'. This means that a single anchor-carrier (see Section 3.2.3 for definition) is used separately for uplink and downlink. Later, in the evaluation of a dual-carrier operation in Section 6.6, a secondary carrier i.e. a non-anchor is added, each in the uplink and downlink.
3. For deciding the configuration of the scheduling scheme, NPDCCH and NPRACH, the approach was as follows:
    i. An initial arbitrary configuration of the scheduler, NPDCCH and NPRACH was considered. Further, an urban environment and in-band deployment mode was assumed.
    ii. A unilateral sensitivity analysis under different network loads, similar to what will be presented later in Section 6.5, of each of the above three configuration components was done. Note that, for the NPRACH configuration, only the resource configuration and the maximum number of random access (RA) attempts were separately varied.

iii. From the above sensitivity analysis, the optimal setting, obtained for majority of the loads, for each component, was combined to derive the final baseline configuration shown in the table.

4. With respect to the NPDCCH configuration (see Section 3.4.1 for parameter definitions), an overlapping configuration (see Figure 3-11) is considered, to allow for more flexibility in varying the parameters $\mathbf{R_{max}}$ and **T**.
5. With respect to the NPRACH configuration (see Section 3.5.2 for parameter definitions), a frequency-multiplexed (see Figure 3-16) resource configuration is used. This was found to perform better than a time-multiplexed configuration because of the flexibility to assign a much higher number of RAOs per second for coverage level 1, where the expected arrival rate is the highest of the three coverage levels.

Table 6-1. Baseline network configuration.

Parameter		Settings		
Coupling loss thresholds per CL		CL 1: [0, 130] dB CL 2: (130, 140] dB CL 3: (140, ∞) dB		
Carrier operation		Single-carrier		
Scheduling scheme		EDDF-SPTF with MGA		
NPDCCH resource configuration		**CL 1**	**CL 2**	**CL 3**
R_{max}		8		
Offset (α)		0		
Periodicity parameter (G)		1.5		
Period (T) (ms)		12		
NPRACH configuration		**CL 1**	**CL 2**	**CL 3**
Resource configuration	**Period (ms)**	80	160	320
	Number of preamble repetitions[11]	2	4	32
	Starting subframe (ms)	8		
	Number of preambles	24	12	12
Maximum number of RA attempts		19	5	7
Backoff interval (ms)		[0, 1024]		
RAR window size (ms)		10*T		
MSG 4 window size (ms)		64*T		

For deriving the baseline configuration, the unilaterally derived optimal settings for the different configuration components were combined, starting from an arbitrary configuration. Note that this baseline configuration so derived is not necessarily globally optimal, due to inter-component effects. In general, such an approach is chosen as a heuristic to arrive at a non-arbitrary baseline for further analysis. Furthermore, this approach can be used to reach the global optimum for a given load, provided several iterations are carried out. In the sensitivity analysis presented in Section 6.5, one such iteration is performed, at the end of which a configuration that is potentially close to the global optimum for all loads is obtained. Henceforth, this obtained configuration will be termed as a 'quasi-optimal' configuration.

[11] The number of preamble repetitions for the three coverage levels have been set based on a minimum target detection probability of 95 % (as per the equation in Appendix A) and assuming a maximum coupling loss of 130, 140 and 154 dB respectively.

6.4 Baseline scenarios: results and analysis

In this section, we will evaluate the baseline scenarios under different loads. The load is determined by the number of smart meter UEs attempting to send power outage notifications, which in turn depends on the extent of the power outage. We model the extent of power outage through a so-called 'outage percentage' which is the percentage of substations covered in the radio cell that is affected by the power outage. Such a power outage scenario models power outages in the high-voltage transmission lines that may result in one or more substations being affected. For the analysis of the baseline scenarios, three outage percentages are considered: 10%, 50% and 100%.

The outline of this section is as follows. In Section 6.4.1, we will look at how the loads compare between different environments for each of the outage percentages. In Section 6.4.2, the simulation results for the baseline scenarios are presented. Finally, a summary is provided in Section 6.4.3.

6.4.1 Comparison of loads for different environments

Figure 6-1 shows, for each environment, in the radio cell under simulation, the number of covered substations, the average number of smart meters per substation and the number of affected smart meters for the three outage percentages. The average number of smart meters per substation is determined by dividing the total number of smart meters covered by the radio cell over the number of covered substations. Note that the plots remain the same irrespective of the deployment mode since the best serving radio cell for each dropped UE is determined from the propagation conditions (coupling loss) which depend only on the environment.

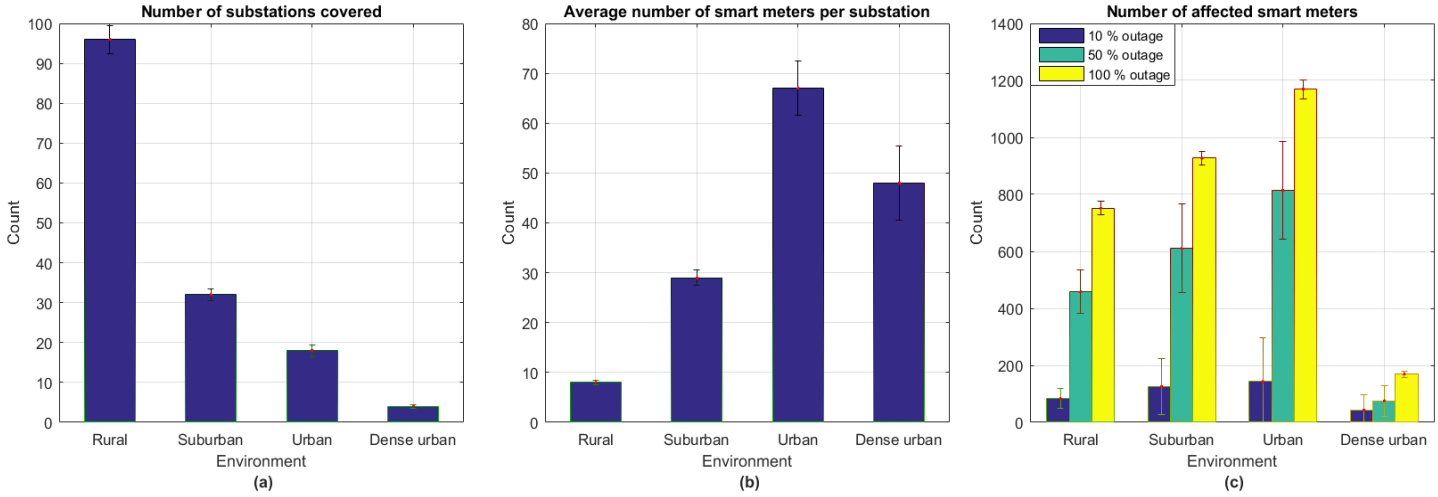

Figure 6-1. Plots showing, for each environment, (a) the number of substations covered in the radio cell, (b) the average number of smart meters per substation and (c) the number of smart meters affected by different outage percentages. The bar values and error bars indicate the mean and standard deviation, respectively, of the obtained values across 100 simulation snapshots.

From the figure, we can see that, although the number of substations per radio cell (plot (a)) decreases from rural towards dense urban, the load due to the number of affected smart meters (plot (c)) shows a different pattern. The network load is highest in urban, followed by suburban, rural and dense urban. This is because of the relative differences in the average number of smart meters per substation between the different environments (shown in plot (b)). Note that the number of affected smart meters can be derived roughly as the product of the number of affected substations (based on the outage percentage) and the average number of smart meters per substation. For example, the number of affected smart meters in the rural environment for 100% outage can determined as the product of 95 (i.e. the value from plot (a)) and 9 (i.e. the value from plot (b)) i.e. 760, which corresponds roughly to the observed value in plot (c).

6.4.2 Comparison of KPIs for different environments and deployment modes

Figure 6-2 shows a comparison of the reliability performance for different environments under an in-band deployment mode. The performance of the other deployment modes is discussed at the end of this subsection.

The main observations from the figure are as follows:

1. The worst reliability performance is observed in the urban environment, for all outage percentages. This is because the load in the urban environment is the highest, as seen in Figure 6-1, for any given outage percentage, which in turn mainly impacts the success rate (shown later in this subsection).
2. Using the baseline configuration, 100% reliability i.e. an optimal performance is achieved in the rural and dense urban environments, for all outage percentages. On the other hand, in the suburban environment, since the load is higher, the reliability decreases to 90% in the 100% outage case. Note that, although the given configuration results in similar performance for most outage percentages, in rural, suburban and dense urban environments, this may not be true for any configuration, in general, because of the difference in non-trivial propagation effects and the load values (see Figure 6-1). In a similar way, an optimal configuration, for a given environment and load, may not necessarily be optimal for the other environments.

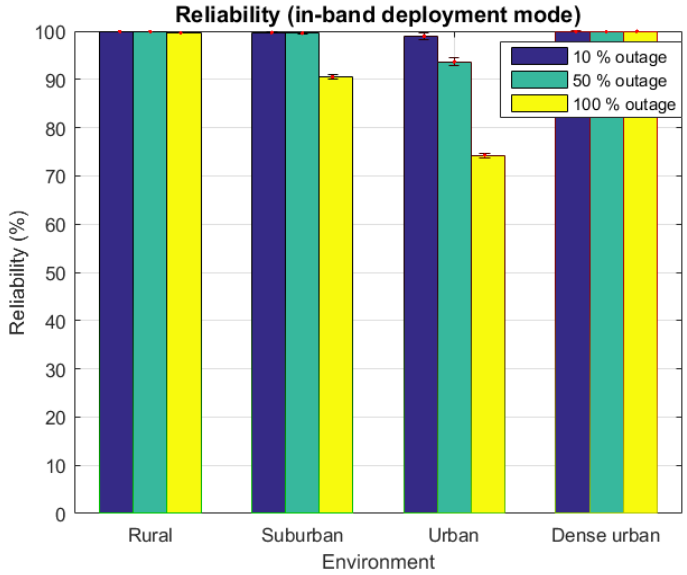

Figure 6-2. Comparison of the reliability performance for each environment under an in-band deployment mode. The error bars indicate the 95 % confidence intervals of the mean value obtained across 100 simulation snapshots.

Since the reliability is derived using the success rate and transfer delay of the successfully delivered power outage notifications (see Equation (6.1)), we next look at these constituent KPIs, to better understand the reliability observations. Figure 6-3 shows a comparison of the success rate and the 95th transfer delay percentile of the successful notifications, for the different environments.

The main observations from the figure are as follows:

1. In general, the success rates and the 95th transfer delay percentile are the worst for the urban environment, which explains the corresponding reliability performance seen above.
2. For suburban and urban environments, at 100 % outage, the success rates are higher than the corresponding reliability values, which means that a certain proportion of successful notifications are not received within the 20 second transfer delay target. This is also partly reflected in the 95th transfer delay percentile observations for 100% outage, where the values are slightly above 20 seconds.

3. The decrease in success rate with increase in load, particularly for the urban environment, is mainly due to preamble collisions and timeouts i.e. a failure to schedule the RAR/MSG 4 within the respective time windows.
4. Except for the dense urban environment, where the load is low, for all other environments, we see a general increase in the 95th transfer delay percentile with an increase in the load. We know from Figure 5-1 that the transfer delay consists of two components i.e. the *access delay* and the *UL data transfer delay*. It is expected that both these components would increase as the load increases, as explained below:
    i. With a higher load, the chances of preamble collisions and/or timeouts increase, thereby increasing the number of attempts that may be required by a UE for a successful random access. With an increase in the number of RA attempts, the *access delay* naturally increases.
    ii. With a higher load, the waiting time for UEs to be scheduled in the UL data transmission phase (see Figure 5-1) would increase, thereby increasing the *UL data transfer* delay.

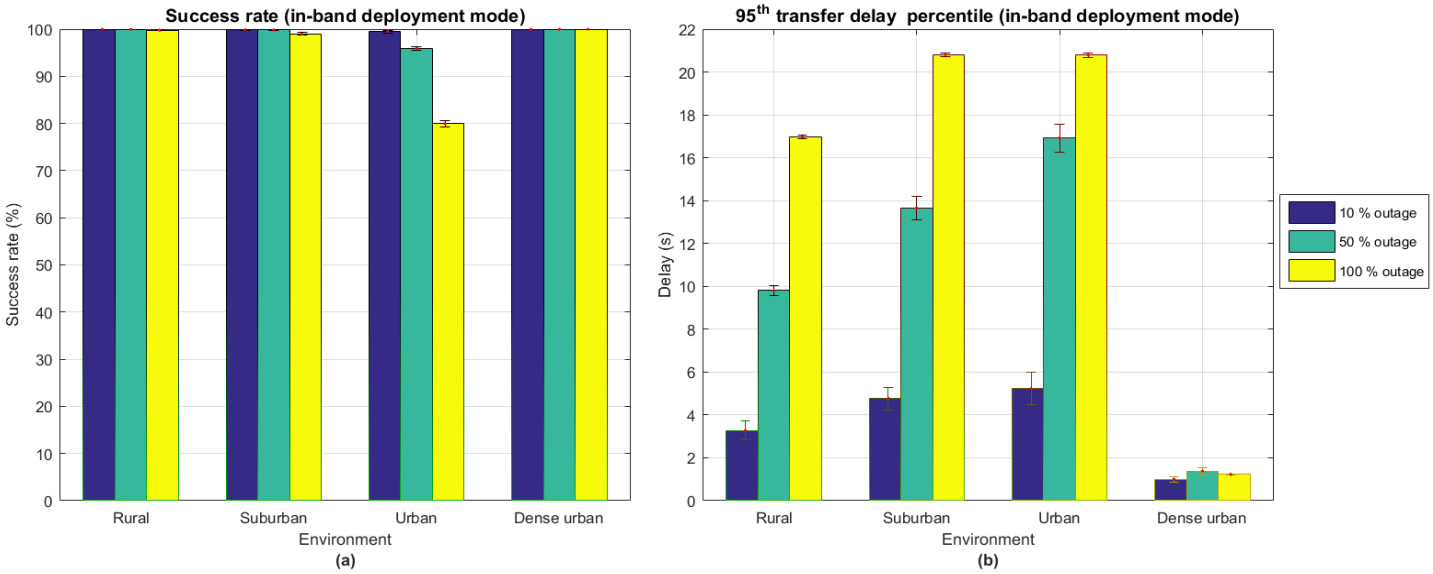

Figure 6-3. Comparison, for different environments in an in-band deployment mode, of (a) the success rate and (b) the 95th transfer delay percentile. The error bars indicate the 95 % confidence intervals of the mean value obtained across 100 simulation snapshots.

So far, we analysed the KPIs for the baseline scenarios with the in-band deployment mode. Upon changing the deployment mode to guard-band or stand-alone, no significant change in the above KPIs was observed (see Appendix F). Although better downlink throughput performance is expected for guard-band and stand-alone modes, as explained in Section 3.2.4, it is not expected to produce a significant gain in the overall performance for the considered scenarios, since the overall data transmission process involves only two downlink message transfers (for RAR and MSG 4) with small message sizes (< 30 bytes).

6.4.3 Summary

In this section, we evaluated the reliability performance of the baseline network configuration in different environments and NB-IoT deployment modes. The performance was evaluated for different outage percentages which determined the load in terms of the number of UEs attempting to send a power outage notification.

The key results from this section can be summarised as follows:

1. For a given outage percentage, the observed load in the considered environments increased in the order: dense urban, rural, suburban and urban, because of the relative differences in the number of

substations covered in the radio cell and the average number of smart meters connected per substation.

2. The worst reliability performance, amongst the considered environments, was observed for the urban environment mainly because the associated load was the highest for any given outage percentage. In the rural and dense urban environments, optimal performance (100% reliability) was achieved with the considered baseline configuration at all outage percentages.
3. The overall reliability performance for all the considered environments was observed to be insensitive to a change in the NB-IoT deployment mode because of the insignificant amount of downlink data involved in the data transmission process.

Due to scope limitations, only a single environment-deployment mode combination is considered for the sensitivity analysis presented in the next section. To this extent, the urban environment is chosen since the same has been considered by 3GPP in [5] as a reference (typical) scenario for NB-IoT deployment. Noting that the performance across the different deployment modes was roughly similar, the in-band mode is chosen since it is expected to be the most likely initial deployment choice for NB-IoT operators as predicted in [51].

6.5 Sensitivity analysis for optimisation

In this section, we will perform a sensitivity analysis of the KPIs to certain components of the baseline configuration. This will help to illustrate the significance of these components in ensuring an optimal configuration and to determine a final configuration of these components which is potentially close to a globally optimal configuration or in other words, quasi-optimal. The sensitivity of the KPIs is evaluated with respect to the following components:

1. NPRACH resource configuration (in terms of the NPRACH period)
2. Maximum number of RA attempts
3. Scheduling scheme
4. NPDCCH resource configuration (in terms of the parameters $\mathbf{R_{max}}$ and **T**)

The optimisation of the above components can be viewed as ‘strategies’ to optimise the network performance using the available resources, which is the focus of this study. A few additional outage percentages, namely 20%, 35% and 70%, are considered for this analysis, to get a more detailed insight on the performance impact due to the variation of load. Figure 6-4 shows the corresponding number of affected smart meters, in the urban environment, for all the considered outage percentages.

Note that in the figure, the number of affected smart meters increases non-linearly with the outage percentage since the smart meters under radio cell coverage are not uniformly spread across the substation cells (see Figure 4-7).

The outline of this section is as follows. In Section 6.5.1 - 6.5.4, the sensitivity analysis of the four components mentioned above is presented, including an overview of the respective optimisation trade-offs. Finally, the derivation of the quasi-optimal configurations is discussed in Section 6.5.5.

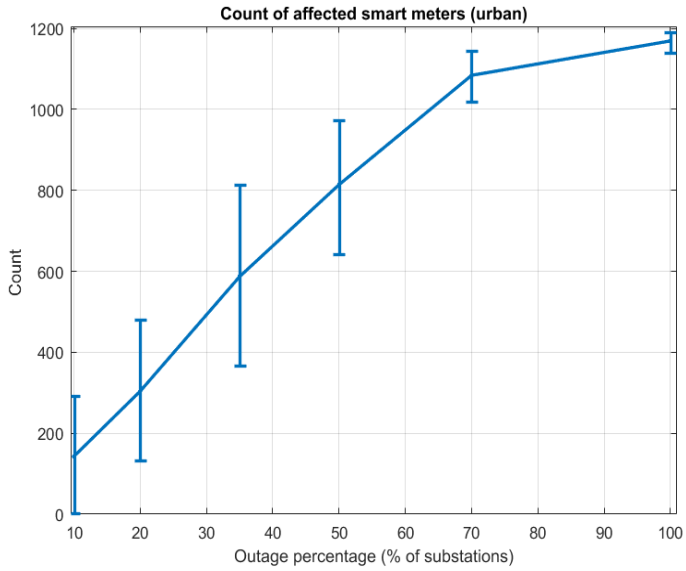

Figure 6-4. Count of affected smart meters, in the urban environment, for the outage percentages considered in the sensitivity analysis. The data points and error bars indicate the mean and standard deviation, respectively, of the obtained values across 100 simulation snapshots.

6.5.1 Impact of the NPRACH resource configuration

6.5.1.1 Overview

Optimisation of the random access channel resources is typically considered as a basic strategy to improve the probability of successful random access in scenarios involving massive network access [68]- [70]. In this analysis, this is illustrated by varying the NPRACH resource allocation for a coverage level, in terms of the number of RAOs per second. With reference to the system description in Figure 5-1, the resources need to be allocated in such a way that there is an optimal trade-off between the preamble detection success probability $\mathbf{p_s\,(t)}$ is and the timeout failure rate $\boldsymbol{\mu_{f1}}$ **(t)**. Such a trade-off exists since the NPRACH and the NPUSCH share the same UL resources, as depicted in Figure 3-16. Thus, the possible gain in success rate achieved by increasing $\mathbf{p_s\,(t)}$, e.g. by increasing the NPRACH resources, may get negated by an increase in $\boldsymbol{\mu_{f1}}$ **(t)** due to a decrease in the NPUSCH resources. Furthermore, fewer NPUSCH resources may also increase the UL data transfer delay and possibly the transfer delay as well.

6.5.1.2 Configuration scenarios

Table 6-2 shows the four evaluated NPRACH resource configurations including that of the baseline configuration in Table 6-1, represented here by Config 2.

From Config 1 → 4, the number of RAOs per second are increased, only for coverage level 1, by decreasing the corresponding NPRACH period by one step (see Section 3.5.2 for the list of possible values). The NPRACH period of 40 ms in Config 1 is the minimum configurable period considering the preamble duration for coverage level 1. The last column in the table depicts a UL resource occupancy metric which indicates the percentage occupancy (the formula used is shown in Appendix F.2) of the total UL time-frequency resources for the NPRACH. The remaining percentage of resources is available for NPUSCH. This metric is thus a quantitative indicator of the allocation of the shared UL resources amongst the NPRACH and NPUSCH. As expected, the UL resource occupancy of the NPRACH increases, though not linearly, with an increase in the number of RAOs per second i.e. the NPRACH resources.

Table 6-2. Summary of the evaluated NPRACH resource configurations.

Configuration name	Period (ms) (CL 1)	RAOs per second (CL 1)	UL resource occupancy of the NPRACH
Config 1	40	600	32%
Config 2 (baseline)	80	300	25%
Config 3	160	150	21%
Config 4	320	75	19%

6.5.1.3 Results

Figure 6-5 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile among the NPRACH resource configurations considered in Table 6-2.

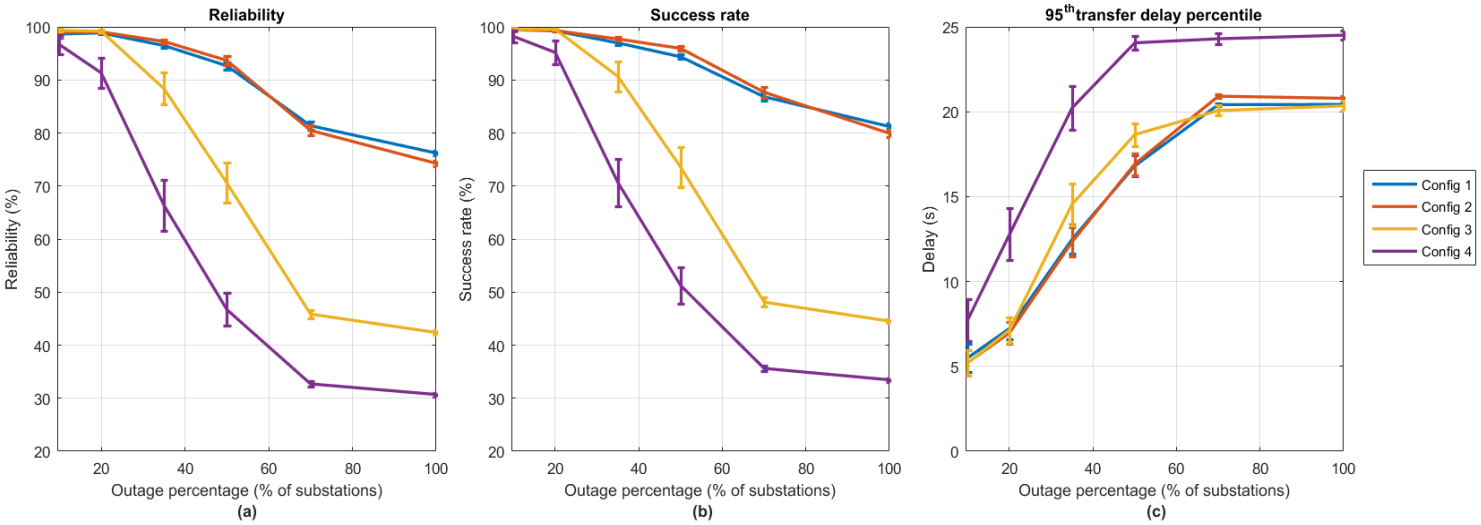

Figure 6-5. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among the considered NPRACH resource configurations.

The main observations from the figure are as follows:

1. Both Config 1 and Config 2 can be considered practically optimal since they show very similar performance, with Config 1 performing marginally better above 70% outage. Furthermore, the optimality of both the configurations is quite robust to the variation in the load.
2. The reliability performance worsens as the UL resource occupancy of the NPRACH is decreased beyond that in Config 2, as is evident from the performance of Config 3 and Config 4.
3. The success rate of Config 1 being slightly lower than Config 2, for low loads is due to a higher timeout failure rate arising from fewer NPUSCH resources, as stated in the overview. At higher loads, the effect of transfer failures due to preamble collisions becomes a dominating factor. This effect is lower for a configuration with a higher number of RAOs per second, which is why the Config 1 and Config 2 plots appear to converge and cross.
4. In Config 3 and Config 4, the number of RAOs per second is too low such that the collision probability is significant, even at low loads, thus resulting in a lower success rate and reliability compared with Config 1 and Config 2.

5. For most loads, we see an increase in the 95th transfer delay percentile with decrease in the success rate. In general, from Config 1/2 → 4, the success probability in an RA attempt by a UE decreases, resulting in a higher total number of RA attempts made by a successful UE. Consequently, the access delay also increases which may be compensated by a decrease in the UL data transfer delay due to an increase in the NPUSCH resources. This is possibly why there is not much difference in the 95th transfer delay percentile values between Config 1 – 3 above 60% outage. In case of Config 4, the resulting access delay is too high to be compensated by the decrease in the UL data transfer delay.

Note that in Figure 6-5(c), the 95th transfer delay percentile for Config 1 and Config 2 slightly decreases beyond 70% outage. This is primarily due to the relative differences in the congestion of the NPRACH resources in the three coverage levels. As per Table 6-1, coverage level 3 has the lowest NPRACH period and hence the lowest number of RAOs per second. Consequently, the success probability of UEs attempting random access, particularly after a coverage level update, in coverage level 3, drops much faster compared to lower coverage levels. With increasing load, depending on the variation in the ratio of the number of UEs succeeding at each coverage level, which would vary across configurations, the contribution to the 95th transfer delay percentile may come from UEs (originally in coverage level 1) updating only upto coverage level 2, thus with a lower number of RA attempts and transfer delay, resulting in a decrease in the percentile value. In few configurations shown in later analyses, this decrease is more prominent

6.5.1.4 Summary

In this section, we evaluated the impact of varying the number of RAOs per second assigned for a coverage level. Increasing the number of RAOs per second meant increasing the NPRACH resources and conversely decreasing the NPUSCH resources. Thus, the aim was to allocate the NPRACH/NPUSCH resources in such a way that there is an optimal trade-off between the achieved preamble collision probability and the timeout failure rate.

The key results obtained from the simulations can be summarised as follows:

1. The baseline configuration and a configuration with higher NPRACH resources resulted in a similar optimal performance, whereas the configurations with lower NPRACH resources than the baseline performed much worse. Furthermore, the optimality of the former two configurations was observed to be quite robust to the variation in the load.
2. The above result was observed because the amount of NPRACH resources assigned in the optimal configurations were high enough such that the preamble collision probability and the timeout failure rate were optimally balanced at practically all loads. In the suboptimal configurations, the assigned NPRACH resources were too low such that collision probability was significantly high.

6.5.2 Impact of the maximum number of RA attempts

6.5.2.1 Overview

The general expectation is that with an increase in the maximum number of RA attempts, the likelihood of preamble collisions, for a UE, in all its attempts would decrease, thereby increasing the success rate. It must be noted that this is not trivial since, with an increase in the allowed number of RA attempts, there may be an increased number of UEs reattempting preamble transmissions, thereby increasing the arrival rate of reattempts $\boldsymbol{\lambda_R(t)}$, as per Figure 5-1. This could, in principle, decrease the preamble detection success probability $\boldsymbol{p_s(t)}$, for a given $\boldsymbol{\lambda(t)}$ and negatively impact the success rate. However, if the $\boldsymbol{\lambda(t)}$ is low enough on an average, an increased number of RA attempts, each after a random backoff time, increases the likelihood that a UE, at some point in time, encounters only a few contending UEs such that its preamble is successfully detected.

Since the underlying idea is to increase the random access success probability by allowing the UE to make a higher number of RA attempts, the trade-off, naturally, is an increase in the UE's access delay and

consequently also the transfer delay. Thus, it is possible that an increase in the success rate may actually result in a decrease in the reliability, because of the violation of the transfer delay target.

6.5.2.2 Configuration scenarios

Table 6-3 shows the four evaluated configurations with different maximum number of RA attempts. From Config 1→ 4, the maximum number of RA attempts for coverage level 1 is increased by four. Config 2 represents the configuration used in the baseline scenarios.

Table 6-3. Summary of the evaluated configurations of the maximum number of RA attempts.

Configuration name	Maximum number of RA attempts (CL 1)
Config 1	15
Config 2 (baseline)	19
Config 3	23
Config 4	27

6.5.2.3 Results

Figure 6-6 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile among the configurations considered in Table 6-3.

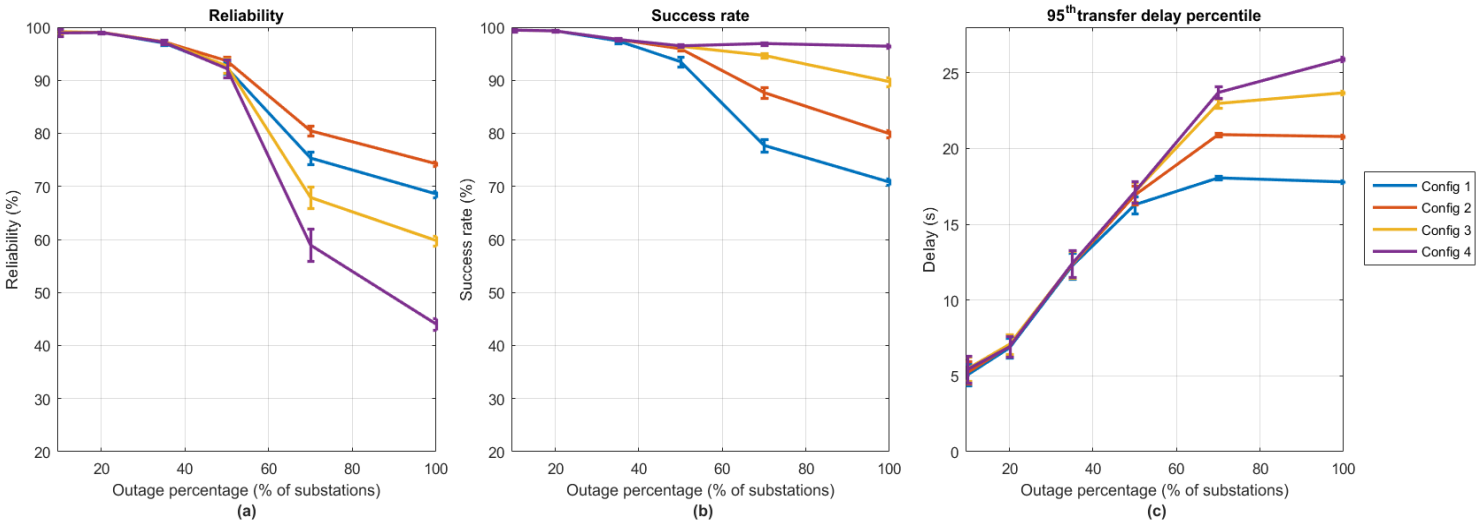

Figure 6-6. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among the considered configurations of the maximum number of RA attempts.

The main observations from the figure are as follows:

1. Config 2 (baseline), with the maximum number of RA attempts as 19, results in the highest reliability performance at all outage percentages.
2. An interesting observation is that the reliability performance does not continuously increase with an increase in the maximum number of RA attempts from Config 1 → 4. We see an initial improvement in reliability from Config 1 → 2, after which the performance worsens as we move from Config 2→ 4.
3. The above observation is, as stated in the overview, due to the net effect of an increase in both the success rate and the 95th transfer delay percentile, as observed for an increase in the maximum number of RA attempts from Config 1 → 4. Particularly, the 95th transfer delay percentile for Config 3 and Config 4 is significantly higher than 20 seconds, the transfer delay target. The net effect is the degradation in the reliability for these configurations compared with that of Config 1 and Config 2.

6.5.2.4 Summary

In this section, we analysed the impact of varying the maximum number of RA attempts on the KPIs. With a higher number of allowed RA attempts, the expectation was that the collision probability seen by the UE in at least one of its attempts may be low enough such that its preamble detection is successful, thereby improving the success rate. However, this improvement was expected to happen at the cost of a higher 95th transfer delay percentile, which in turn may negatively impact the reliability. The key results obtained from the simulations can be summarised as follows:

1. An improvement in the reliability with an increase in the maximum number of RA attempts was observed upto a certain maximum value. Beyond that, the reliability was found to decrease. Consequently, the baseline configuration showed the highest reliability performance for all loads.
2. The above observations were due to the net effect of the success rate-transfer delay trade-off i.e. an increase in the maximum number of RA attempts resulted in an increase in both the success rate and the 95th transfer delay percentile.

6.5.3 Impact of the scheduling scheme

6.5.3.1 Overview

Scheduling is expected to have an impact in the last two phases depicted in Figure 5-1, i.e. the RAR/MSG 4 and UL data transmission phases. The schedulers proposed in Chapter 5 are listed below:

1. EDDF with MGA
2. SPTF with MGA
3. EDDF-SPTF with MGA
4. EDDF-SPTF with MMA
5. EDDF-SPTF with LGA

The use of due date-prioritised scheduling combined with different UL subcarrier allocation schemes, is aimed at maximising the number of UEs that can be served within their respective time windows i.e. to maximise mainly the service rate $\mu_{s1}(t)$ for the available set of physical resources and given load $p_s(t) * \lambda_t(t)$. For a given $p_s(t) * \lambda_t(t)$, a higher $\mu_{s1}(t)$ is expected to result in an improvement in the success rate. Furthermore, maximising $\mu_{s1}(t)$ would help to minimise the total number of RA attempts required by a UE before successful access, thus minimising the access delay. However, since UEs in the last two phases contend for the same physical resources, depending on the applied prioritisation scheme, the cost of increasing $\mu_{s1}(t)$ could be a lower $\mu_s(t)$, due to a higher waiting time of UEs in the UL data transmission phase. This could result in a higher UL data transfer delay in general. This may be acceptable, though, if there is a significant reduction in the access delay, which would lead to an overall lower transfer delay. As stated in Chapter 5, the choice of prioritisation schemes involves a trade-off of varying degree between assigning priority to a UE's due date and priority to its transmission time. Similarly, the UL subcarrier allocation schemes involve a trade-off between the scheduling capacity and the achieved bit rate for high-SINR users.

6.5.3.2 Configuration scenarios

Table 6-4 shows a summary of the evaluated configurations with the different proposed schedulers. Each scheduler is represented by a combination of a prioritisation and UL subcarrier allocation scheme. Config 3 represents the configuration used in the baseline scenarios.

Table 6-4. Summary of the evaluated configurations with the different proposed schedulers, each a combination of a prioritisation and a UL subcarrier allocation scheme.

Configuration name	Prioritisation scheme	UL subcarrier allocation scheme
Config 1	EDDF	MGA
Config 2	SPTF	MGA
Config 3 (baseline)	EDDF-SPTF	MGA
Config 4	EDDF-SPTF	MMA
Config 5	EDDF-SPTF	LGA

6.5.3.3 Results

Figure 6-7 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile among the schedulers considered in Table 6-4.

The main observations from the figure are as follows:

1. Config 3, with the EDDF-SPTF and MGA scheduler, practically results in the highest reliability at all outage percentages. Thus, the combination of the two traditional EDDF and SPTF principles is indeed more effective than their stand-alone forms. Furthermore, since Config 1 performs much better than Config 2, it shows that the EDDF principle plays a more significant role in achieving optimal performance, compared to the SPTF principle.
2. Since the transmission time of a UE is expected to have a greater impact on other UEs, at high loads than at low loads, Config 1 (with only EDDF prioritisation) performs increasingly worse than Config 3 (with the added SPTF prioritisation) as the load increases.
3. For the same prioritisation scheme (EDDF-SPTF), the reliability performance shows an improvement with an increase in the UL subcarrier allocation granularity, as is evident when comparing Config 3 - 5. Note that this granularity increases in the order: LGA, MMA and MGA.
4. The above observation is due to an increase and decrease in the success rate and the 95th transfer delay percentile respectively, with an increase in the UL subcarrier allocation granularity. This shows that a decrease in granularity increases the waiting time of UEs, resulting in an increase in the timeout failure rate, the access delays and ultimately the 95th transfer delay percentile. It must be noted that the small packet sizes involved also have a role to play in achieving optimal performance despite the lowering of the bit rates with maximisation of subcarrier allocation granularity.

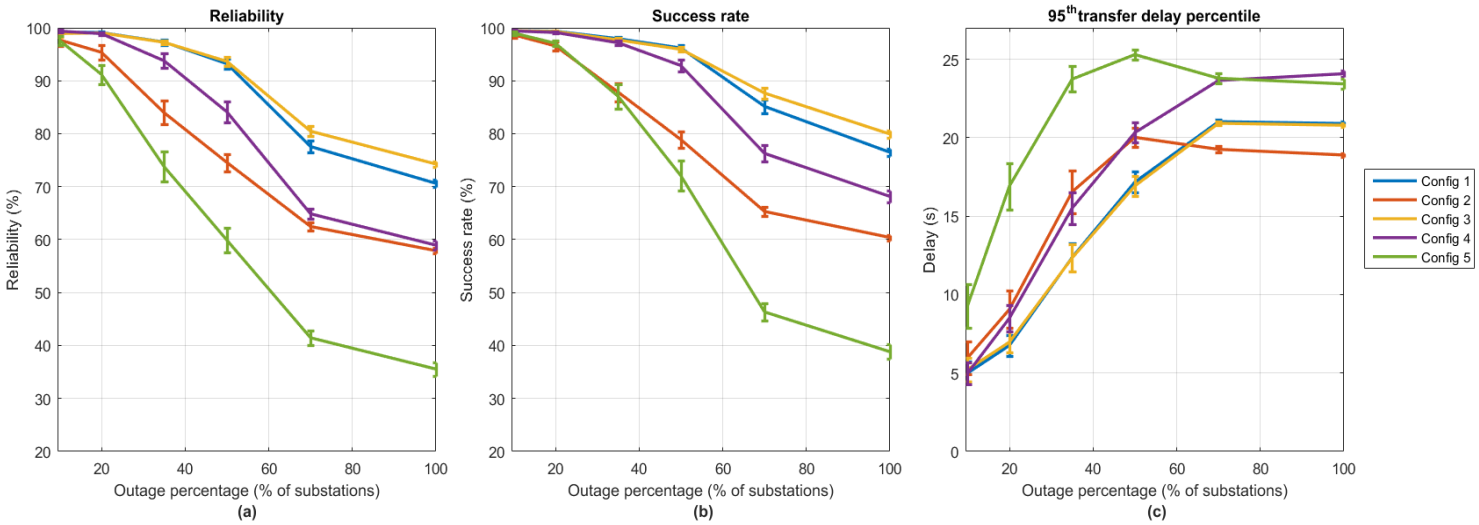

Figure 6-7. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among the considered schedulers.

Note that the decrease in the 95th transfer delay percentile for Config 2 and Config 5 beyond 50% outage is due to the same reason mentioned in Section 6.5.1.3.

The presence of the EDDF component in the prioritisation scheme, in general, leads to a higher UL data transfer delay as compared to, say, using the SPTF principle. This is because the UEs in the UL data transmission phase, with a delay budget in the order of seconds, are likely to be assigned least priority in the presence of UEs in the RAR/MSG 4 transmission phase, with a delay budget in the order of milliseconds. This impacts the reliability performance, particularly at high loads. For example, in the figure, the reliability above 50% outage for Config 3, is roughly 5-8 % less than the success rate value, which is an indicator of the proportion of successful UEs not meeting the transfer delay target. On the other hand, for the SPTF-based scheduler i.e. Config 2, this difference is smaller even though with Config 2, UEs are expected to have a higher access delay, as indicated by the lower achieved success rate.

A 'crude' solution to reduce the UL data transfer delay and the possible impact on the reliability is to offload the UL data transmission phase UEs to a different physical carrier, as is achieved with a dual-carrier operation, discussed in Section 6.6.

6.5.3.4 Summary

In this section, we evaluated the sensitivity of the KPIs to the scheduling scheme. By optimising the scheduler, the reliability could be maximised by minimising the timeout failure rate and thereby maximising the success rate and minimising at least the access delay. The schedulers proposed in Chapter 5 aimed to achieve this using a combination of a prioritisation and a UL subcarrier allocation scheme. The key simulation results are summarised as follows:

1. The EDDF-SPTF with MGA scheduler achieved the highest reliability for practically all outage percentages, thereby performing better than the schedulers which used the traditional EDDF and SPTF principles. Furthermore, the EDDF principle was observed to have a greater impact in achieving optimal performance as compared to the SPTF principle whose effect was prominent only at high loads.
2. The reliability performance improved with an increase in the granularity of the UL subcarrier allocation i.e. from LGA → MGA. With small packet sizes involved, this was mainly due to a decrease in the waiting time of UEs which resulted in an improvement in both the success rate and 95th transfer delay percentile.

6.5.4 Impact of the NPDCCH resource configuration

6.5.4.1 Overview

Optimisation of the NPDCCH resources involves adjusting the NPDCCH resource allocation, in terms of the parameters $\mathbf{R_{max}}$ and **T** (see Section 3.4.1), which influences the number of scheduling opportunities (SOs) per second ($\mathbf{R_{max}}$*1000/**T**) for the given coverage level. The resources need to be allocated in such a way that there is an optimal trade-off between the waiting time of UEs in the given coverage level and those in the other coverage levels. As described in Section 3.4.1, this trade-off is relevant in an 'overlapping' NPDCCH configuration (see Figure 3-11) considered in this analysis. If the number of SOs per second for a given coverage level is too high for the given load, then it may block a high number of UEs of other coverage levels from being scheduled, resulting in their high number of timeout failures. Conversely, if the number of SOs per second is too low, then the waiting time of UEs in the given coverage level will be high, particularly at high loads, resulting in their high number of timeout failures and an overall suboptimal performance.

6.5.4.2 Configuration scenarios

Table 6-5 shows a summary of the evaluated NPDCCH resource configurations, each represented by a specific value of parameters $\mathbf{R_{max}}$ and **T**, indicated in the second and third columns. The fourth column indicates the number of SOs per second for the corresponding configuration. From Config 1 → 3, the number of SOs per second is increased by increasing the $\mathbf{R_{max}}$ by one step (see Section 3.4.1 for the list of possible values) for a fixed value of **T**. From Config 3 → 4, an increase in the number of SOs per second is only possible with an increase in **T** as well. Additionally, Config 4 represents the maximum configurable number of SOs per second.

Table 6-5. Summary of the evaluated NPDCCH resource configurations with different R_{max} and T values.

Configuration name	R_{max} (CL 1)	T (ms) (CL 1)	SOs per second (CL 1)
Config 1	1	8	125
Config 2	2	8	250
Config 3	4	8	500
Config 4 (baseline)	8	12	667

6.5.4.3 Results

Figure 6-8 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile among the NPDCCH resource configurations considered in Table 6-5.

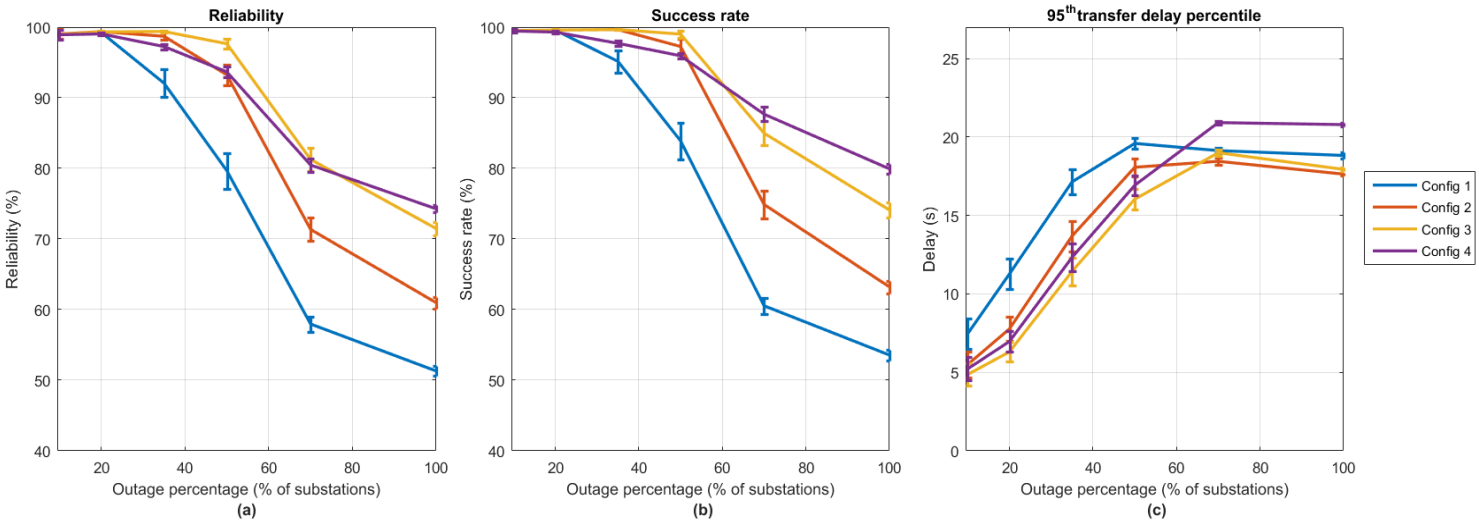

Figure 6-8. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among the considered NPDCCH resource configurations.

The main observations from the figure are as follows:

1. A load-dependent optimal configuration is observed. Config 3 has the highest reliability upto roughly 75% outage beyond which Config 4 (baseline) shows the best performance, though only slightly higher than Config 3. From this point on, we see a clear increase in the reliability performance with increase in the number of SOs per second, i.e. from Config 1 → 4.
2. The success rate plot shows a somewhat similar trend as seen for the reliability. The slightly better success rate of Config 1 and Config 2 as compared with Config 3 at outage percentages less than 45% is due to lesser blocking of NPDCCH resources caused to UEs of other coverage levels, thereby resulting in lesser number of timeout failures of those UEs. However, as the load increases further, the waiting time of UEs in coverage level 1, where the expected load is the highest, becomes a dominant factor, leading to a higher number of timeout failures for these UEs in configurations with a lower number of SOs per second. Thus, at high loads, Config 3 achieves the highest success rate.
3. Config 1 shows worse reliability and success rate performance, for most loads, compared to the other configurations because the allocated number of SOs per second is too low.
4. The general observations of the success rate are analogous to the corresponding observations of Figure 6-5, where different NPRACH resource configurations were compared. It was seen that a less than the maximal allocation of the number of RAOs per second (analogous to the number of SOs per second) showed the highest success rate at low loads whereas a maximal allocation appeared to result in a slightly better performance at high loads. Further, the performance was much worse if the allocated number of RAOs per second were too low (similar to Config 1 here).

5. For most loads, the 95th transfer delay percentile shows a decrease with increase in the number of SOs per second, for a fixed **T**, i.e. from Config 1 to 3, because of a decrease in the waiting time for scheduling. For a higher **T**, such as in Config 4, the timeout of RAR/MSG 4 windows would take longer since the window size is a multiple of **T** (see Table 6-1). Thus, for a comparable total number of failed RA attempts, a UE in Config 4 may experience a higher access delay than a UE in the other configurations. This possibly explains why Config 4 has a higher 95th transfer delay percentile compared to Config 3.

6.5.4.4 Summary

In this section, we evaluated the sensitivity of the KPIs to the NPDCCH resource configuration. The parameters $\mathbf{R_{max}}$ and **T**, for coverage level 1, were varied, to arrive at a configuration in which there is an optimal trade-off between the between the waiting time of UEs in the given coverage level and those in the other coverage levels. The key results from the simulations can be summarised as follows:

1. A load-dependent optimal configuration was observed wherein the configuration with a maximal number of SOs per second showed the best performance at high loads whereas, at low loads, a configuration with less than maximal number of SOs per second was optimal.
2. The above result was mainly due to the success rate being majorly impacted, at high loads, by the waiting time of UEs (lesser with higher number of SOs per second) in the given coverage level and, at low loads, by the blocking of SOs (lesser with lower number of SOs per second) for UEs in the other coverage levels.

6.5.5 Derivation of the quasi-optimal configuration

In the previous four subsections, we looked at the impact of alternative configurations that unilaterally deviated from the baseline with respect to a specific component (dimension) such as the maximum number of RA attempts and the scheduling scheme. In the case of the NPRACH resource configuration, a particular configuration showed marginally better performance than the baseline at high loads. Similarly, in the case of the NPDCCH resource configuration, we saw that two alternative configurations performed better than the at low loads. Based on this, a set of configurations can be identified which, for a specific range of loads, show the best reliability performance amongst all the configurations which performed better than or similar to the baseline configuration. It is possible that these load-specific best-performing configurations, in all, deviate from the baseline in more than one component. In such a scenario, it would be useful to evaluate the performance of a configuration that 'combines' some or all of these differing components, with the expectation that it may have the best performance at all the considered loads.

Thus, we have two types of candidate quasi-optimal configurations:

1. Load-specific best-performing configuration.
2. 'Combined' configuration.

After comparing all these candidate configurations, we can conclude on the quasi-optimal configuration for the entire range of considered loads.

6.5.5.1 Configuration scenarios

The candidate quasi-optimal configurations considered in this analysis are summarised in Table 6-6. Config 1 (differing in the NPDCCH configuration from the baseline), derived from Table 6-5, is the configuration which performed best amongst all configurations throughout the sensitivity analysis, upto roughly 70% outage.

Config 2 (differing in the NPRACH period from the baseline), derived from Table 6-2, on the other hand, performed the best beyond 70% outage. Although the performance of this configuration was only marginally better than the baseline, it has been chosen here to evaluate the performance of a 'combined' configuration i.e. which is Config 3. Config 3 is thus composed of the differing NPDCCH and NPRACH components from Config 1 and Config 2 respectively.

Table 6-6. Summary of the evaluated candidate quasi-optimal configurations. All parameters, where applicable, correspond to the configuration for coverage level 1.

Configuration name	NPRACH period (ms)	Maximum number of RA attempts	Scheduler	NPDCCH configuration (R_{max}, T)
Config 1 (<=70% outage)	80	19	EDDF-SPTF with MGA	(4,8)
Config 2 (> 70% outage)	40	19	EDDF-SPTF with MGA	(8,12)
Config 3 (combined)	40	19	EDDF-SPTF with MGA	(4,8)

6.5.5.2 Results

Figure 6-9 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile between the candidate quasi-optimal configurations in Table 6-6. We can see that Config 1 and Config 3 have practically the highest reliability performance upto roughly 50% outage. Beyond this, Config 3 clearly performs the best. Note that Config 1 and Config 3 differ in their NPRACH configuration. Contrary to the observations in Section 6.5.1, here the performance, at high loads, of the configuration with higher NPRACH resources is significantly better due to the different NPDCCH resource configuration. Likewise, the reliability and success rate plots of Config 2 and Config 3, which differ in their NPDCCH resource configuration, appear to converge and cross at a certain load, though beyond the considered range. In general, this is similar to the observations in Section 6.5.4.

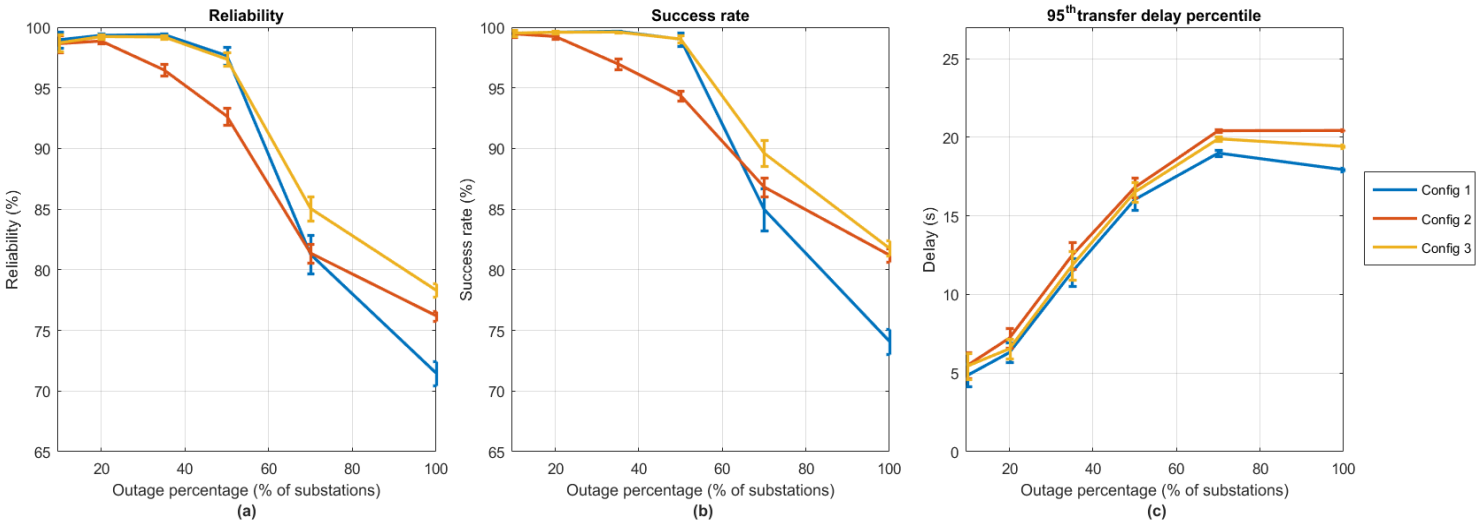

Figure 6-9. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among the candidate quasi-optimal configurations.

The above results indicate that a single configuration may not, in principle, achieve optimal performance at all loads. However, for the range of loads the study is restricted to, Config 3 can be considered as the quasi-optimal configuration. It can also be noted that the achieved reliability below 50% outage, with Config 3, is close to 100%. Thus, there is room for further significant performance improvement only above 50% outage. Though not covered in this study, this may be investigated through another iteration of sensitivity analysis with Config 3 as the baseline.

6.5.5.3 Summary

In this subsection, the objective was to derive a quasi-optimal configuration for the entire range of loads considered in this study, for an urban environment. To this extent, three candidate configurations were selected, based on the sensitivity analysis performed previously. From the simulation results, a quasi-optimal configuration for the entire range of considered loads was identified, with the relevant configuration component settings listed in Table 6-7. With this configuration, at least 98% reliability was achieved for outage percentages upto 50%.

Table 6-7. Obtained quasi-optimal configuration settings.

Configuration	**Setting**		
Scheduler	EDDF-SPTF (prioritisation) with MGA (UL subcarrier allocation)		
NPDCCH resource	**CL 1**	**CL 2**	**CL 3**
R_{max}	4	8	8
T	8	12	12
NPRACH	**CL 1**	**CL 2**	**CL 3**
NPRACH period	40	160	320
Maximum number of RA attempts	19	5	7

6.6 Evaluation of a dual-carrier operation

6.6.1 Overview

In the previous section, we looked at different ways in which the available resources of a carrier can be optimised to maximise the reliability. In this section, we will investigate the performance gain achieved by the addition of physical resources, through a dual-carrier operation which is a specific case of the multi-carrier operation described in Section 3.2.3. In the dual-carrier operation, a non-anchor carrier is configured, each in the uplink and the downlink in addition to the existing anchor carrier pair. Note that such a resource expansion is a last resort in case a further improvement in the performance of the optimal configuration is needed.

The additional non-anchor carrier in UL and DL can be solely used for scheduling and data transmission purposes i.e. without any overhead of NPSS/NSSS/NPBCH/NPRACH. A UE can be configured to listen to the non-anchor carrier after MSG 4 is successfully received [71]. Thus, UEs in the UL data transmission phase, as per Figure 5-1, can be independently scheduled in a separate physical carrier, thereby reducing the contention for scheduling of physical resources between the sets of UEs in the last two phases. This can reduce the waiting time, compared to a single-carrier operation, for UEs in the RAR/MSG 4 scheduling phase and improve the success rate and access delay via a reduction in the timeout failure rate, $\boldsymbol{\mu_{f1}(t)}$. Further, in the last phase, the general waiting time (and hence the UL transfer delay) is lower and in the absence of overheads (NPSS/NSSS/NPBCH), more DL resources, mainly in the time domain, are available for NPDCCH/NPDSCH.

The performance gain with the addition of a non-anchor carrier may not be as significant as that achieved by adding resources to the random access channel to address its congestion, e.g. by adding an anchor carrier. Further, for the in-band mode, the downlink PSD boosting possible for a non-anchor carrier is lower (3 dB) compared to the anchor carrier (6 dB) [72]. However, the key advantage in adding a non-anchor carrier is that it can be done dynamically, when the need for additional data channel resources arises, such as during a power outage event. An anchor-carrier, in contrast, would need to be active throughout, even when the UEs are idle, to provide signalling information. Consequently, in an in-band mode, adding a non-anchor carrier would have lesser impact on the LTE capacity as compared to adding an anchor carrier.

6.6.2 Results

Figure 6-10 shows a comparison of the reliability, the success rate and the 95th transfer delay percentile between a single-carrier and dual-carrier operation. In both cases, the rest of the configuration corresponds to the baseline configuration in Table 6-1.

The main observations from the figure are as follows:

1. We can see that there an improvement of about 5% at high loads, in the reliability, with a dual-carrier operation, while at low loads, particularly below 35% outage, there is not much improvement since a single carrier apparently provides sufficient resources to achieve the highest reliability performance.
2. The improvement in reliability is mainly due to the reduction, of 2-3 seconds, in the 95th transfer delay percentile, given that there is only a marginal increase observed in the success rate. As stated in the overview, a reduction in the transfer delay is expected because of a decrease in both the access and UL data transfer delays.
3. The marginal improvement in success rate indicates that the impact to the RAR/MSG 4 UEs is low even in the single-carrier operation. This is due to the EDDF-SPTF scheduling scheme, which tends to assign the least priority to UEs in the UL data transfer phase because of their higher delay budgets (order of seconds) compared to that of the RAR/MSG 4 UEs (order of ms). Consequently, it can be stated that the reduction in the 95th transfer delay percentile in the dual-carrier operation is mainly due to the reduction in the UL data transfer delay rather than the access delay.

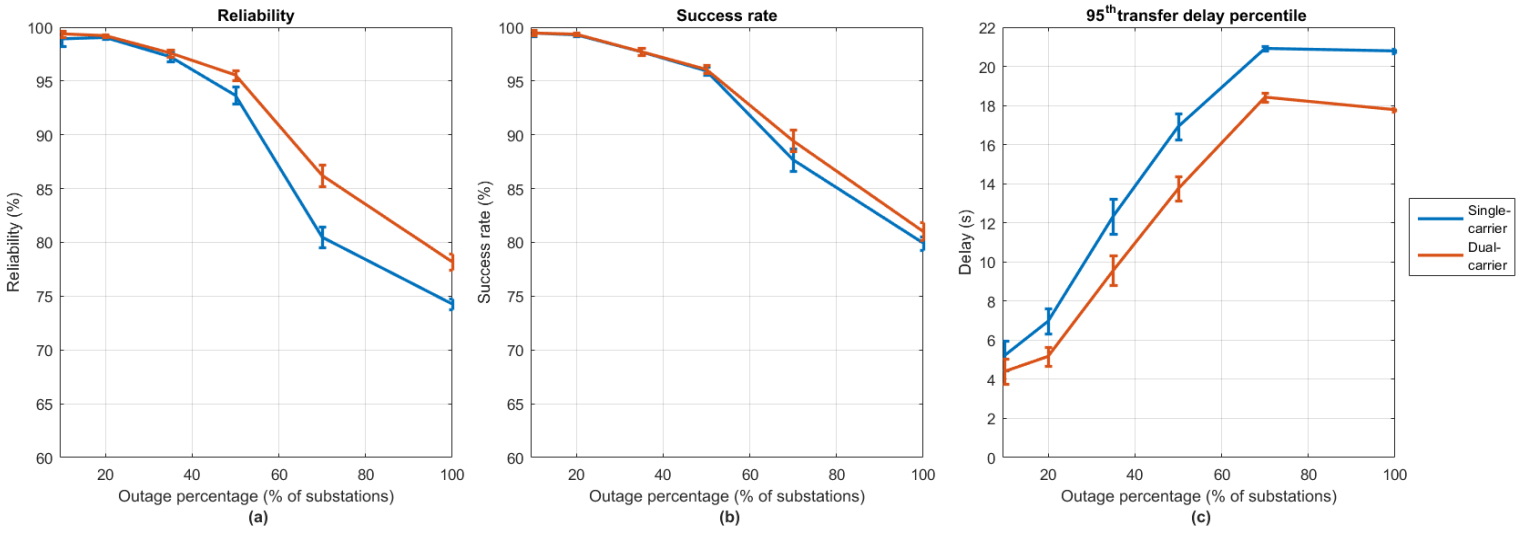

Figure 6-10. Comparison of (a) the reliability, (b) the success rate and (c) the 95th transfer delay percentile among a single-carrier and dual-carrier operation.

Based on the observations of the 95th transfer delay percentile, the dual-carrier operation can be viewed as a 'crude' solution to reduce the UL data transfer delay in a single-carrier operation, particularly with the optimal EDDF-SPTF scheduling scheme as was discussed in Section 6.5.3. Further, with the reduction in the 95th transfer delay percentile, there is further room for optimisation, e.g. a further increase in the maximum number of RA attempts.

6.6.3 Summary

In this section, we compared the performance of a single-carrier and a dual-carrier operation (with one non-anchor carrier pair). With a dual-carrier operation, since the UL data transmission occurs on a separate physical carrier, it was expected that the timeout failure rate of the RAR/MSG 4 UEs and the waiting time of the UL data transfer UEs would reduce, resulting in an improvement in both the success rate and the 95th transfer delay percentile, and therefore the reliability. The results indeed showed this, with the reliability improvement mainly resulting from a significant decrease in the 95th transfer delay percentile. Although, the

reliability improvement with a non-anchor carrier is not as significant as that may be achieved through addition of an anchor carrier, the former is still a more efficient option considering it can be activated dynamically if needed during a power outage event.

6.7 Summary

In this chapter, we evaluated the performance of NB-IoT in terms of the achieved reliability, in the context of the ORM use case in smart grids. The evaluation was done under different network configuration, deployment and power outage scenarios. First, we looked at the performance in different environments (rural, suburban, urban and dense urban) and NB-IoT deployment modes (in-band, guard-band and stand-alone). Then, a set of strategies were evaluated to maximise the reliability performance either by optimising the available physical resources of a carrier or by addition of physical resources (dual-carrier operation). The focus in this study was on the former set of strategies.

For the chosen baseline configuration, we saw that the urban environment had the worst performance mainly because its network load for a given outage percentage was the highest amongst the considered environments. Further, the reliability performance was observed to be insensitive to a change in the NB-IoT deployment mode i.e. the in-band, guard-band and stand-alone modes, mainly because of the insignificant amount of downlink data involved in overall data transmission process.

Four components of the network configuration were considered in the optimisation study, listed as follows:

1. The NPRACH resource configuration (in terms of the NPRACH period)
2. The maximum number of RA attempts
3. The scheduling scheme
4. The NPDCCH resource configuration (in terms of the parameters $\mathbf{R_{max}}$ and **T**)

Considering an urban environment and an in-band deployment mode, the sensitivity of the reliability performance to a unilateral change in each of the above components was analysed, based on which a final quasi-optimal configuration was derived. Table 6-8 shows a summary of the main underlying trade-offs involved in the optimisation of these components.

Table 6-8. Summary of the four optimisation components analysed in this study.

Optimisation component	Underlying trade-offs	Comments
NPRACH resource configuration (NPRACH period)	Preamble collision probability (↑↓) vs timeout failure rate (↑↓)	Optimal setting was same at all loads
Maximum number of RA attempts	Success rate (↑↓) vs transfer delay (↑↓)	Optimal setting was same at all loads
Scheduling scheme	• Granularity (↑↓) vs bit rate for high-SINR UEs (↓↑) • Priority to due-date (↑↓) vs priority to transmission time (↓↑)	Optimal setting was same at all loads
NPDCCH resource configuration ($\mathbf{R_{max}}$, **T**)	Waiting time of UEs of given CL (↑↓) vs waiting time of UEs of other CLs (↓↑)	Optimal setting was different at low and high loads

In the *optimisation of the NPRACH resource configuration* for a coverage level, the trade-off between the preamble collision probability and the timeout failure rate was exploited, which exists because the NPRACH and the NPUSCH share the same uplink resources. Overall the configuration with maximal number of RAOs per second was observed to be quite robust in its optimality at all loads, which was also noted during the derivation of the quasi-optimal configuration. At the same time, at low loads, a configuration with less-than-maximal NPRACH resource allocation was observed to perform equally well.

In the *optimisation of the maximum number of RA attempts*, the trade-off between the success rate and the transfer delay was exploited. As a result of this trade-off, the optimal setting was seen to be limited by the increase in the transfer delay above the target value.

In the *optimisation of the scheduling scheme*, the combined trade-offs in the choice of the prioritisation schemes and UL subcarrier allocation schemes were exploited. In the choice of the UL subcarrier allocation schemes, the trade-off between the subcarrier allocation granularity and the bit rate for high-SINR UEs was exploited, whereas in the three flavours of the prioritisation scheme, the trade-off between assigning priority to a UE's date and priority to its transmission time was exploited. Overall, the EDDF-SPTF with MGA scheduler was found to be the optimal scheduler at all loads.

In the *optimisation of the NPDCCH resource configuration,* the parameters $\mathbf{R_{max}}$ and **T** of the NPDCCH configuration were varied. The key trade-off exploited was between the waiting time of UEs in the given coverage level and those in the other coverage levels. Additionally, it was seen that a certain configuration was optimal only for a range of loads. A maximal allocation of the number of SOs per second showed optimal performance at high loads whereas a slightly lesser allocation was optimal at low loads.

Based on the sensitivity analyses, a set of candidate quasi-optimal configurations were identified. From the simulation results, the configuration shown in Table 6-7 was concluded to be the desired quasi-optimal configuration for the considered range of loads in the urban environment. With this configuration, at least 98% reliability was achieved for outage percentages upto 50% which corresponds to roughly 800 affected smart meters/households in the radio cell, on an average.

Finally, we looked at the performance gain achieved by means of the addition of physical resources i.e. through a *dual-carrier operation*. With the given configuration, only upto a 5% increase in reliability was achieved. Although the gain with the addition of a non-anchor carrier is not as significant as may be achieved through addition of an anchor-carrier, the former may still be a preferred option since it can be configured dynamically and has lesser impact on the LTE capacity in an in-band mode. Further, with the achieved reduction in the 95^{th} transfer delay percentile, it can be used as a 'crude' solution to reduce the UL data transfer delay in a single-carrier operation and enable further improvement of the reliability performance.

Chapter 7. Conclusions and future work

In this thesis, a comprehensive performance evaluation of NB-IoT (3GPP Release 13) was done, in the context of the Outage Restoration and Management (ORM) use case in smart grids. NB-IoT is a relevant massive IoT technology solution for providing connectivity to smart meters within the Advanced Metering Infrastructure (AMI) in smart grids, mainly because it enables low cost and complexity of devices, supports coverage enhancement techniques, has broad industry support and provides flexible deployment options for operators. ORM, which involves the smart meter notifying the utility in the event of a power outage or restoration, is an important use case for utilities as it aids in accurate power outage localisation and fast recovery, depending on the achieved reliability performance in the network. The reliability metric indicates the percentage of the total generated power outage notifications received successfully within a transfer delay target. From a network point-of-view, this use case involves massive network access by devices in a short span of time, resulting in a congestion of the physical resources, particularly affecting the random access channel. In such scenarios, maximising reliability becomes a challenge.

The aim of this study was to investigate possible strategies to maximise the reliability performance, particularly through optimisation of the available network resources. To this extent, a realistic simulation model was developed, incorporating the relevant characteristics of the energy distribution and mobile networks, the traffic characteristics of the ORM use case and the relevant PHY and MAC layer aspects of NB-IoT. Using this simulation model, an optimisation study of NB-IoT for ORM was performed, considering an urban environment and in-band deployment scenario, characterised by a household density of 1500/km^2 and a radio network inter-site distance of roughly 1.7 km. A use case like ORM, although relevant in the context of smart meters, was neither considered in the design phase of NB-IoT nor evaluated in previous studies. Hence, this study succeeds in at least providing initial insights on the achievable reliability performance for ORM-based network traffic in NB-IoT. Furthermore, the results and conclusions from this study can also, in principle, be applied to other (potential) use cases with a similar traffic model and a comparable transfer delay target.

The outline of the final chapter is as follows. In Section 7.1, key conclusions from the presented results in the previous chapter are derived, and in Section 7.2, recommendations are provided for possible extensions of the work in this thesis.

7.1 Conclusions

Based on the results presented in Chapter 6, the main conclusions that can be drawn are as follows:

1. **Sensitivity of the reliability performance to network configuration components**
   It can be concluded that the ORM reliability performance is sensitive to at least the following components of the NB-IoT network configuration:
   i. The NPRACH resource configuration
   ii. The maximum number of random access attempts
   iii. The scheduling scheme
   iv. The NPDCCH resource configuration

   Therefore, it is important to optimise the above components to achieve maximal reliability performance for a given outage percentage. For near-optimal reliability performance, the recommended configuration of the above four components is shown again in Table 7-1. Since the design of a suitable NB-IoT scheduler for ORM-based traffic is a key contribution of this study, it can be said that the proposed EDDF-SPTF with MGA scheduler is the recommended option since it showed the best performance amongst all the candidate schedulers.

Although the configuration shown in the table below was derived in the context of the urban environment, it is likely that it will perform optimally or near-optimally for the remaining three environments also, particularly the dense urban and rural environments where the load is much lower for the same outage percentage. At the same time, in the other three environments, certain configurations which performed suboptimally in the urban environment may perform equally well compared to the below configuration, due to the lower load.

Table 7-1. Recommended configuration for near-optimal reliability performance.

Configuration	**Setting**		
Scheduler	EDDF-SPTF (prioritisation) with MGA (UL subcarrier allocation)		
NPDCCH resource	**CL 1**	**CL 2**	**CL 3**
R_{max}	4	8	8
T	8	12	12
NPRACH	**CL 1**	**CL 2**	**CL 3**
NPRACH period	40	160	320
Maximum number of RA attempts	19	5	7

2. **Sensitivity of the optimal configuration settings to the carrier deployment mode**
   NB-IoT offers three modes of deploying a Frequency-Division Duplex (FDD) carrier pair – within the LTE carrier (in-band), in the guard-band of the LTE carrier (guard-band) and in a separate spectrum (stand-alone). The implications of these three modes are primarily in the downlink throughput performance. Given that the amount of downlink data involved in the ORM traffic model is not significant, the expectation was that the performance would be largely insensitive to the deployment mode. The baseline evaluations indeed showed similar performance across deployment modes, regardless of the environment and the outage percentage. Thus, we can conclude that the choice of the NB-IoT deployment mode is highly unlikely to impact the optimal configuration settings.

3. **Sensitivity of the optimal configuration settings to the load**
   It is generally desirable to have a configuration whose optimality is robust, i.e. which performs optimally or near-optimally even if loads or other circumstances (e.g. UE distribution) are different from what is expected. However, the performance of the configuration in Table 7-1 indicated suboptimal performance beyond a certain load, due to the impact of the NPDCCH resource configuration. Even though this load lies beyond the range considered in this study, it is still relevant since practical urban environment household densities and radio cell load could, in principle, be higher than the generic assumptions of this study. Thus, it can be concluded that the optimal configuration settings, with regards to the NPDCCH resource configuration, is not robust against load variations.

4. **Achievable performance and suitability of NB-IoT for ORM**
   With a baseline configuration, nearly 100% reliability was achieved for all outage percentages in the rural and dense urban environments, and for the majority of the considered outage percentages in the suburban environment. For the urban environment, with the configuration in Table 7-1, a reliability performance of 98% or higher is achievable for the majority of the outage percentages. Additionally, the dual-carrier operation supported in NB-IoT can further improve the performance of a single-carrier operation. Overall, it can be concluded that, for the rural and dense urban environments, NB-IoT is a suitable communications technology to facilitate ORM. For the suburban and urban environments, NB-IoT is suitable for most, if not all the outage percentages.

7.2 Future work

This thesis provided initial insights on possible strategies to optimise the NB-IoT network for a use case such as ORM. These insights also lead us towards topics which could be interesting areas for further research, discussed as follows:

1. **Study of the sensitivity of the optimal configuration to different transfer delay targets**
   In this study, a near-optimal configuration was derived assuming a transfer delay target of 20 seconds for ORM, as specified in the literature. However, this does not exclude the possibility that the target could be reduced, in practice, depending upon the needs of the utility. Consequently, one or more of the near-optimal configuration settings will change, e.g. a reduction of the maximum number of RA attempts. Thus, it would be interesting to study, if other configuration settings, e.g. the scheduler, also change with alternative transfer delay targets, e.g. 10 or 15 seconds.

2. **Design of a self-optimisation scheme for NB-IoT**
   In a typical deployment scenario, the NB-IoT network, by default, will be optimised for the regular traffic, e.g. smart metering data. During a power outage event, the network must detect the sudden surge in traffic, estimate the load and dynamically tune the relevant network configuration components towards an appropriate value such as that in Table 7-1. Thus, a logical next step could be to design such a 'self- optimising' scheme for NB-IoT. Although self-optimising schemes proposed by researchers [73], in the context of LTE/LTE-A, may be used as a starting point, applicable limitations in NB-IoT need to be considered in the design. For example, the minimum periodicity with which network configuration changes can be broadcast will influence how quickly the system can respond to load changes.

3. **Study of random access load control strategies for NB-IoT**
   Certain strategies for random access load control have been proposed in the literature which may be assessed in combination with a self-optimising scheme discussed above. An example of such a strategy is Dynamic Access Barring (DAB) [74] where, based on the estimated load and a threshold load level, the network broadcasts a probability factor for UEs of different classes, e.g. emergency services, utilities. Devices attempt access if a random number drawn, between 0 and 1, is less than the probability factor, otherwise they perform a random backoff. Such a method will involve a trade-off between the success rate and the transfer delay, due to the postponement of random access. The challenge would be in optimising the threshold load level and the probability factor such that there is a reliability improvement over the non-DAB scenario.

4. **Evaluation of the suitability of NB-IoT for other domains or smart grid use cases**
   Due to scope limitations, the study in this thesis was limited to smart meter-related use cases within the distribution segment of the power grid. An area for further research is the suitability study of other potential smart grid use cases such as in the categories of Distribution Automation (DA), Distribution Customer Storage (DCS) and Electric Transportation (ET). Furthermore, possible realistic scenarios where a mix of these use cases share the same NB-IoT network resources can be investigated.
   Apart from smart grids, there are also other relevant verticals (see Figure 1-1) such as logistics, tracking and smart agriculture, for massive IoT technologies, which may present different kinds of challenges, e.g. with respect to mobility and battery life. This can be an area for further research as well.

Appendices

A. Decision rules for event-driven state transition

This section describes the decision rules for the event-driven state transition model (Figure 4-3) that is used for simulation of an uplink transmission in NB-IoT.

Let **K** be the index of the current subframe in the downlink and the number of UEs be **N**. Here, **K** is used as a time reference (similar to a Transmission Time Interval (TTI) counter in LTE) for the simulation.

Key steps

1. **K → 1**
2. State of all **UEs → S_0** where **S_0** corresponds to 'No packet in buffer' (in Figure 4-3).
3. For all UEs, set preamble transmission counter **preambTrans → 1**
4. Check the state of all **N** UEs, apply the decision rules in corresponding to the state of each UE and update the state, if applicable.
5. If all non-failed UEs have completed their transmission successfully,

   STOP

   Else

   K → K + 1

   Go to step 4

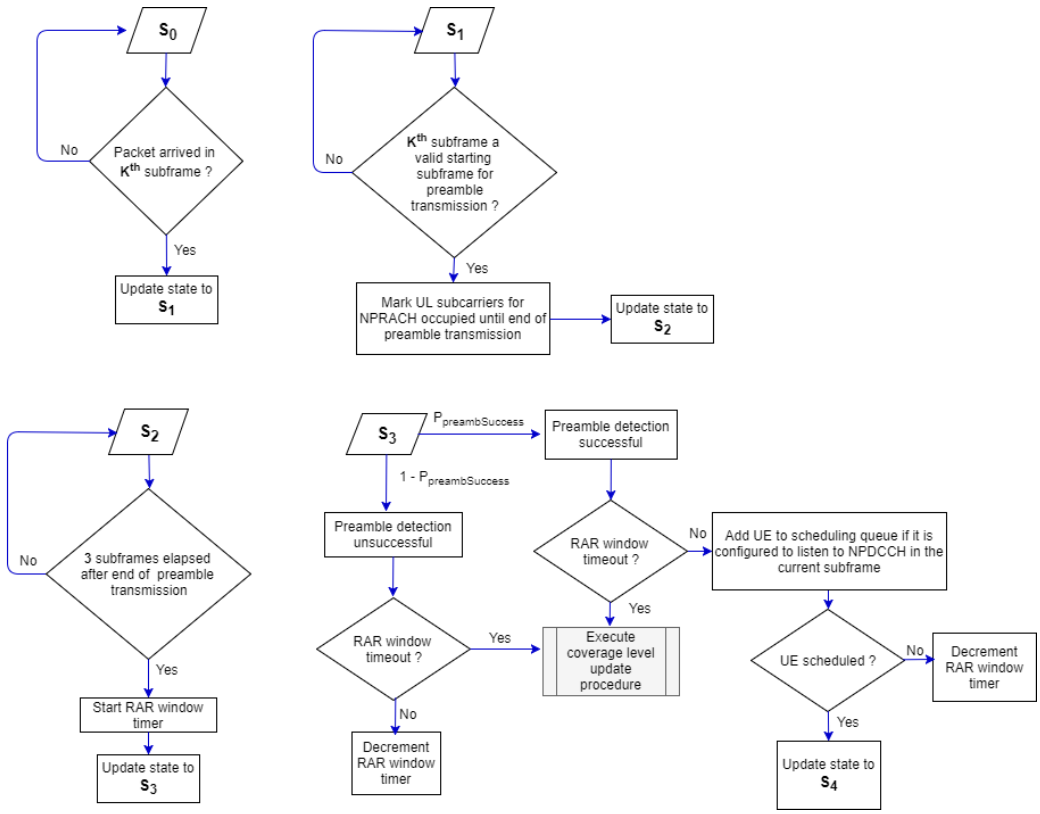

Figure A-1. Decision rules for states 0 to 3.

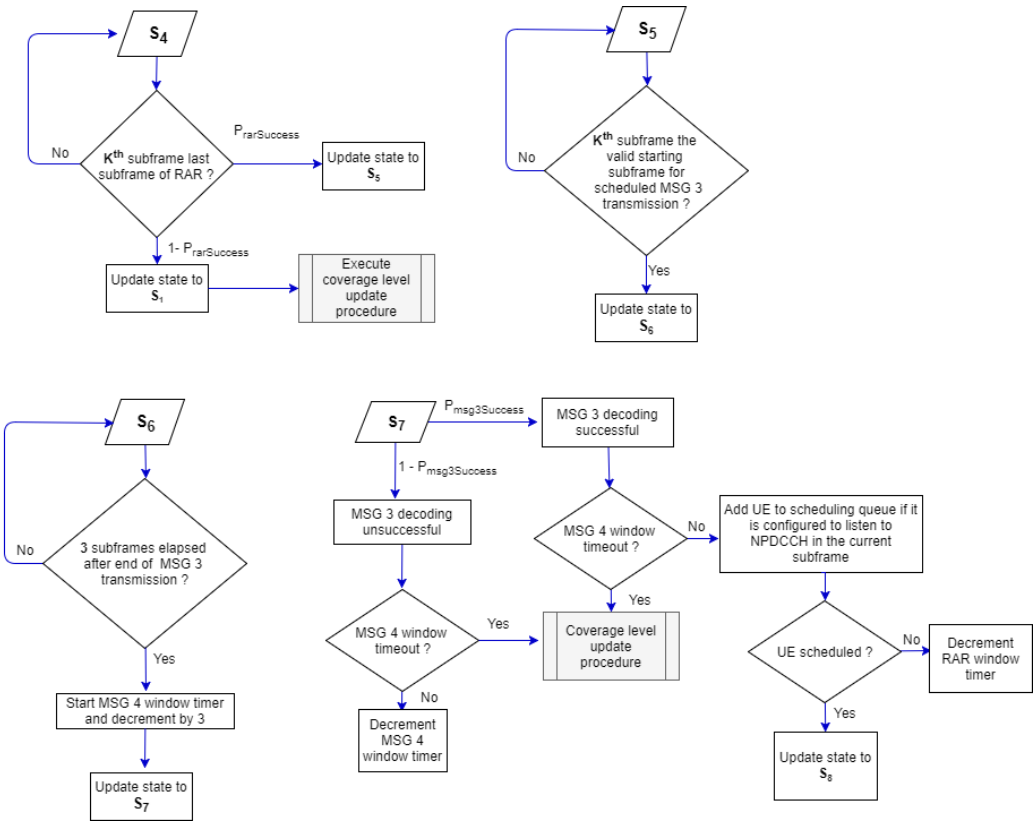

Figure A-2. Decision rules for states 4 to 7.

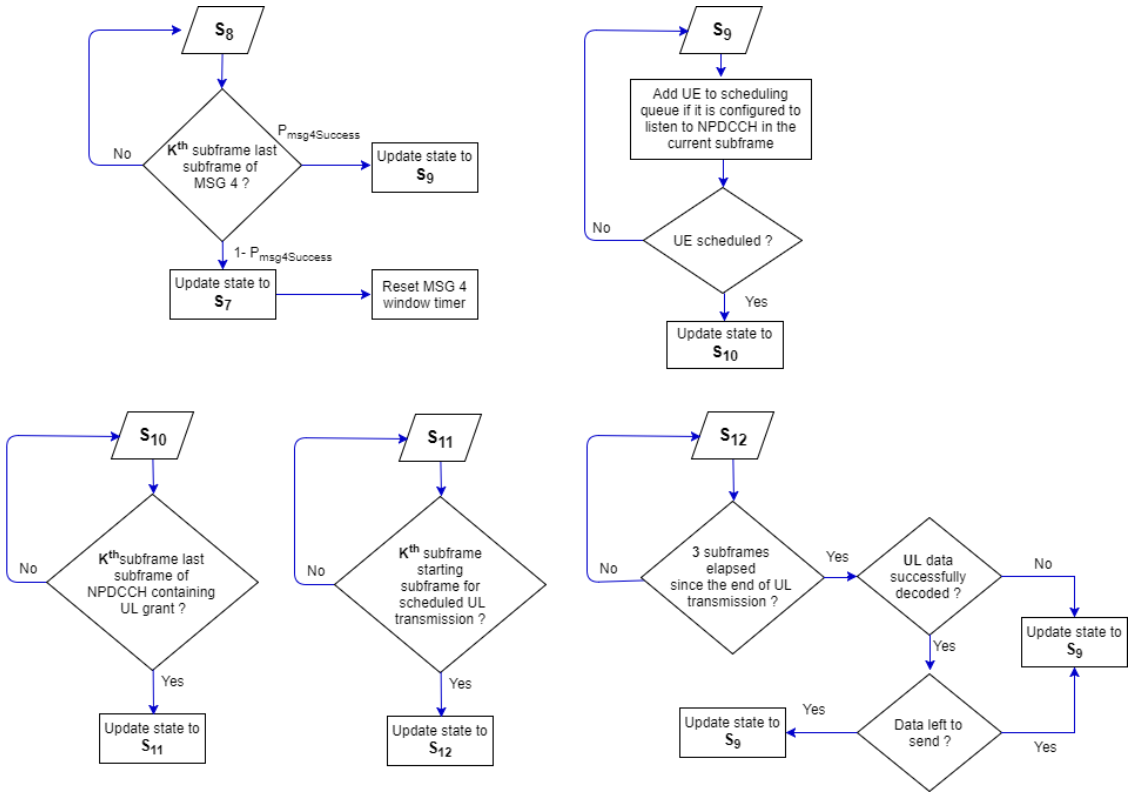

Figure A-3. Decision rules for states 8 – 11.

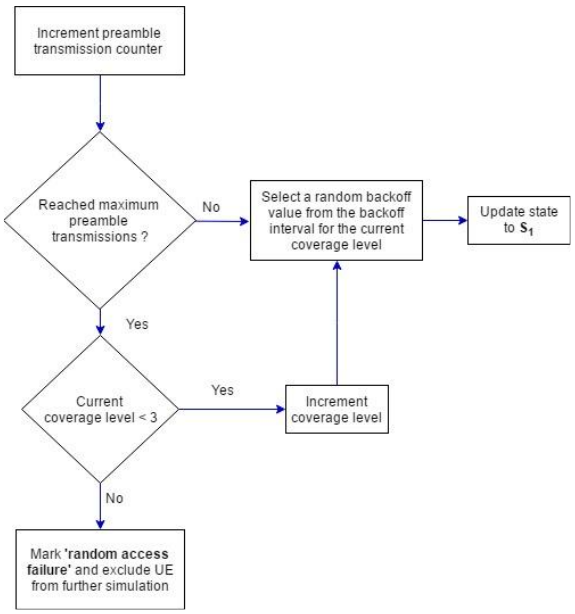

Figure A-4. Coverage level update procedure (for states 3, 4 and 7).

The notations in Figure A-1 to Figure A-4 are defined below:

1. $\mathbf{P_{preambSuccess}}$ is the success probability of detecting a non-collided preamble.
2. $\mathbf{P_{rarSuccess}}$ is the success probability of decoding the NPDSCH transport block containing the RAR.
3. $\mathbf{P_{msg3Success}}$ is the success probability of detecting the NPUSCH transport block containing the RAR.
4. $\mathbf{P_{msg4Success}}$ is the success probability of detecting the NPDSCH transport block containing MSG 4.

The probabilities 2 to 4 above are calculated based on the expected block error rate for the assigned MCS and the DL/UL SINR, derived from the PHY layer model Section 4.3.5. Calculation of $\mathbf{P_{preambSuccess}}$ is explained below.

Calculation of $P_{preambSuccess}$

The analysis which follows is partly adapted from [22].

The overall success probability of the detecting a given preamble that is transmitted only from a single UE i.e. with no collision is the probability that the total received power of the preamble at the base station is above a minimum threshold power. This can be mathematically expressed as follows:

$$P_{\text{preambSuccess}} = 1 - \Pr\ [R_s < T]^{N_{r,CL}} = \left(1 - \left(1 - Q\left(\frac{T-\mu}{\sigma}\right)\right)^{N_{r,CL}}\right) \quad \text{(B.1)}$$

$$\sigma = \sqrt[2]{\sigma_{\text{shadow}}^2 + \sigma_{\text{BPL}}^2} \quad \text{(B.2)}$$

with

- $\mathbf{R_S}$ the received signal power at the base station
- **T** the receiver sensitivity at the base station
- **μ** the mean received power at the base station
- $\mathbf{N_{r,\,CL}}$ "the number of preamble repetitions configured for the coverage level of the UE
- **σ** the cumulative standard deviation of shadowing ($\boldsymbol{\sigma_{shadow}}$) and building penetration loss ($\boldsymbol{\sigma_{BPL}}$)
- **Q (.)** the standard Q function

Note that the parameter **T** above is configured as -120 dBm, the minimum configurable received target power for NPRACH preamble as per [49]

B. Building Penetration Loss (BPL) model assumptions

The Building Penetration Loss (BPL) model is based on an adapted COST 231 NLOS model proposed in [5] and shown in Figure B-1. The BPL parameter assumptions for Dense Urban (DU), Urban (U), Suburban (SU) and Rural (RU) environments are given in Table B-1. The parameter values for urban environment are based on [5], whereas for the remaining environments, they are chosen such that the resulting mean BPL aligns with the mean BPL assumptions for the corresponding environments in [59].

$$BPL = W_e + \max(\Gamma_1, \Gamma_3) - GFH$$

BPL: Building Penetration Loss
W_e : External wall loss (uniformly distributed either in the range $4-11\ dB, 11-19\ dB$ or $19-23\ dB$)

$$\Gamma_1 = W_i * p$$

W_i : Loss in internal walls = $4-10\ dB$ (uniformly distributed)
p : Number of penetrated walls = $0, 1, 2$ or 3

$$\Gamma_3 = \alpha * d$$

α : Penetration distance coefficient = $0.6\ dB/m$
d : Penetration distance = $0-15\ m$(uniformly distributed)

$$GFH = n * G_n$$

n : Floor number = $0, 1, 2, 3$ or 4 (uniformly distributed)
G_n : Floor height gain per floor = $1.5\ dB/floor$

Figure B-1. Model for deriving the BPL [5].

Table B-1. Assumptions for external and internal wall losses.

Assumptions for the external wall penetration loss			
W_e	4 – 11 dB	11 – 19 dB	19 – 23 dB
% of devices uniformly distributed in the range	25 % (DU, U) 45 % (SU) 65 % (RU)	65 % (DU, U) 45 % (SU) 25 % (RU)	10 % (DU, U, SU and RU)
Distribution and assumptions of the internal wall penetration loss			
Percentage of devices mapped to case p = 3 (with remaining devices equally distributed among cases p = 0,1,2)	15 % (DU, U) 5 % (SU, RU)		
Assumption for dependency of internal wall penetration loss	One value of W_i is randomly generated and applied to all internal walls		

C. Calculation of downlink and uplink interference

With reference to Equations (4.2) and (4.3), $\mathbf{I_{eNB}}$ and $\mathbf{I_{UE}}$ are calculated as follows:

$$I_{UE,k} = \sum_{i=1}^{56} \alpha \times P_{tx,eNB} \times CL_{DL,i,k} \qquad (C.1)$$

$$I_{eNB} = \frac{1}{(18x10^3xT)} \sum_{t=1}^{T} \sum_{i=1}^{56} \sum_{k=1}^{N_{UE}} \beta_{t,k} \times P_{tx,UE} \times CL_{UL,i,k} \qquad (C.2)$$

with

- $\mathbf{P_{tx,\,eNB}}$ the transmit power (mW) of the interfering eNodeB **i**
- $\mathbf{P_{tx,\,UE}}$ the transmit power (mW) of UE **k** in the simulation cell
- $\mathbf{CL_{DL,i,k}}$ and $\mathbf{CL_{UL,i,k}}$ the coupling loss, in the downlink and uplink respectively, between the interfering eNodeB **i** and UE **k** in the simulation cell
- $\boldsymbol{\alpha}$ the activity factor of an interfering eNodeB, which refers to the proportion of the total transmit power that contributes towards interference, on an average
- $\boldsymbol{\beta_{t,\,k}}$ the activity flag for UE **k** at time sample **t** (ms). $\boldsymbol{\beta_{t,\,k}}$ is 1 if the UE is active and 0 if not

- N_{UE} the number of UEs in the simulation cell
- **T** the number of time samples over which the average is computed

Equation for $I_{UE,k}$ refers to the sum of average downlink received power from each of the interfering cells. Equation for I_{eNB} refers to the time average of the total uplink interference power per unit bandwidth (Hz) from the simulation cell to all interfering cells. Note that, although I_{eNB} denotes the average interference power at the simulation cell contributed by the uplink traffic in the neighbouring cells, it is calculated based on the uplink traffic pattern of the simulation cell based on the symmetricity assumption, given that the neighbouring cell traffic is not actually simulated. Similarly, the activity factor **α** in Equation (C.1) is determined based on the downlink traffic pattern of the simulation cell.

The key challenge in the above symmetricity assumption is that the downlink and uplink traffic pattern in a given simulation snapshot depends on the settings of **α** and I_{eNB} respectively whose values are, in turn, dependent on the traffic pattern itself. This 'chicken-egg problem' is solved through an iterative approach where initially, arbitrary values of **α** and I_{eNB} are chosen and based on the observed traffic pattern, new values of **α** and I_{eNB} are derived and configured as the new setting for the next iteration. This process is continued until the configured and derived settings nearly converge.

D. Derivation of SINR-MCS mapping for a target BLER of 10 %

The general corrected form of the Shannon capacity equation can be expressed as follows:

$$S_{max}\ (\text{bits/s/Hz}) = BW_{eff}.\,n\,.\log_2(1+\frac{10^{\text{SINR (dB)}}}{SINR_{eff}}) \quad \text{(D.1)}$$

with

- BW_{eff} the system bandwidth efficiency correction
- $SINR_{eff}$ the SINR implementation efficiency correction
- **η** the general correction factor

Uplink

Based on the NPUSCH link level simulation results for an AWGN channel in [61], the corrected Shannon capacity equation for single tone uplink and multi-tone uplink are derived (through curve fitting) as follows:

$$S_{max,\text{single tone}} = 0.36 * \log_2\left(1+10^{\frac{SINR}{10}}\right) \quad \text{(D.2)}$$

$$S_{max,\text{multi tone}} = 0.36 * \log_2\left(1+10^{\frac{SINR+2}{10}}\right) \quad \text{(D.3)}$$

The key difference between single-tone and multi-tone transmissions is the use of differential BPSK/QPSK in the former and non-differential QPSK in the latter. For an AWGN channel, it has been reported in [63] that non-differential QPSK has an approximate 2 dB gain over differential QPSK. Hence, this factor is added in Equation (D.3).

The maximum spectral efficiency (bps/Hz) corresponding to a modulation scheme, subcarrier number, transport block size (TBS) and resource unit (RU) allocation can be calculated as follows:

$$S = \frac{TBS + CRC}{N_{RU} * T_{RU} * SC_{spacing}} \quad \text{(D.4)}$$

with

- **TBS** the Transport Block Size (bits)
- **CRC** the size (bits) of the Cyclic Redundancy Check code. It equals 24 bits for the NPUSCH [48]
- N_{RU} the number of resource units allocated
- T_{RU} the duration (seconds) of a resource unit for the number of subcarriers allocated (1,3,6 or 12)
- $SC_{spacing}$ the subcarrier spacing (Hz) i.e. 15000 Hz or 3750 Hz

Using Equations D.2-4, a mapping can be derived for single-tone and multi-tone allocations between every MCS index and the minimum required SINR for 10 % BLER target. For simplicity in scheduling, the spectral efficiency associated with every MCS index is chosen as the maximum value obtained across all possible RU values.

Downlink

For the downlink, the NPDSCH link level results in [62] are used for deriving the correction factors. The corrected Shannon capacity equation is as follows:

$$S_{max} = 0.65 * \log_2(1 + \frac{10^{\frac{SINR}{10}}}{1.79}) \quad \text{(D.5)}$$

Note that in the downlink, scheduling is always with twelve subcarriers and hence the maximum spectral efficiency is calculated as follows:

$$S = \frac{TBS + CRC}{N_{SF} * 0.001 * 180000} \quad \text{(D.6)}$$

with

- **TBS** the Transport Block Size (bits)
- **CRC** the size (bits) of the Cyclic Redundancy Check code. It equals 24 for the NPDSCH [48]
- **N_{SF}** the number of subframes allocated

A similar mapping between SINR and MCS, as described in the uplink case, is derived for the downlink.

E. SINR-MCS step functions for NB-IoT physical data channels

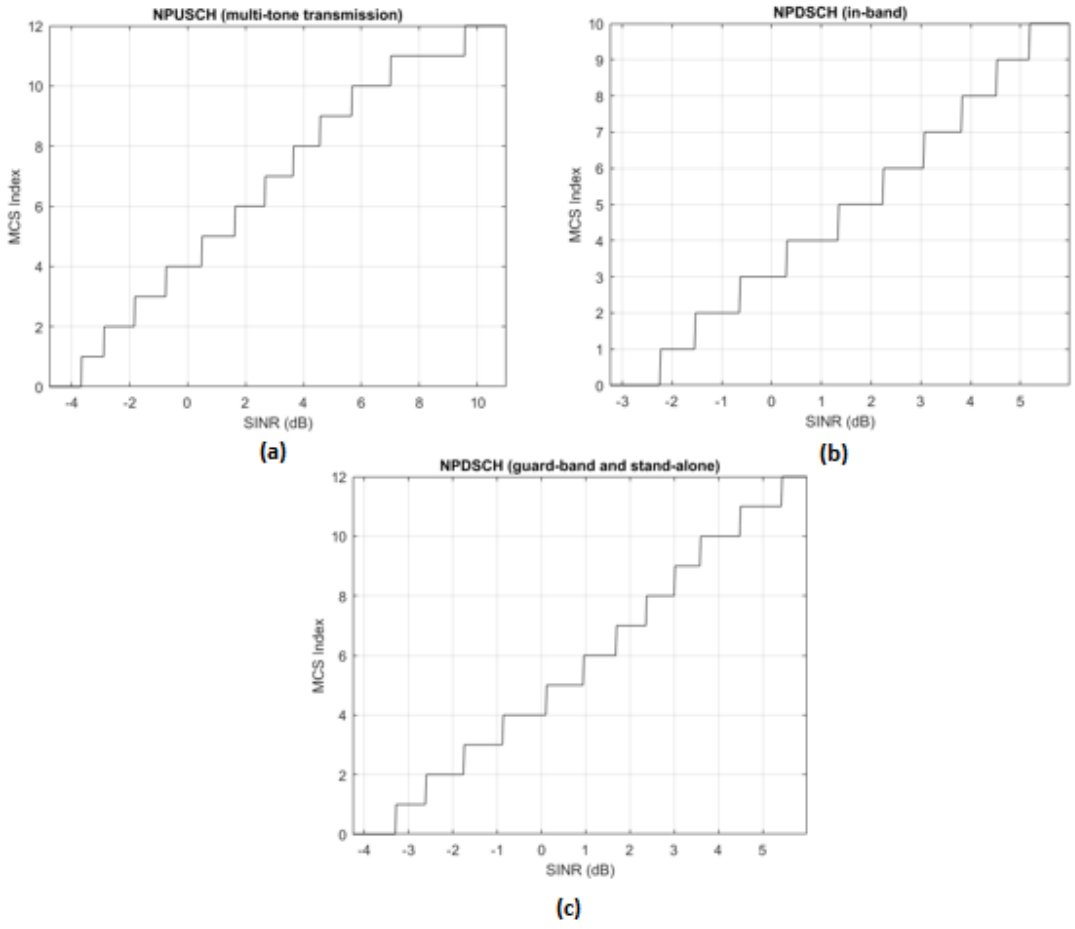

Figure E-1. SINR-MCS step functions for (a) the NPUSCH (multi-tone transmission), (b) the NPDSCH (in-band deployment mode) and (c) the NPDSCH (guard-band and stand-alone deployment modes).

F. Additional simulation results

1. **Reliability (%) performance for guard-band and stand-alone deployment mode**

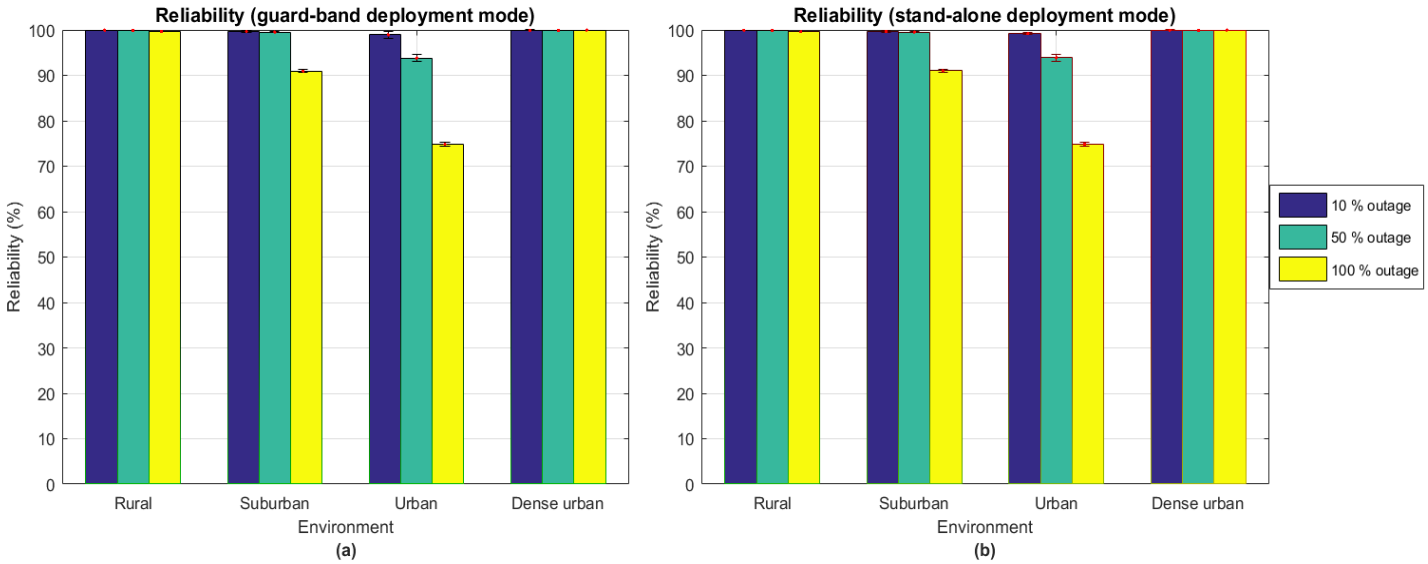

Figure F-1. Comparison of the reliability performance with the baseline configuration for different environments under (a) the guard-band and (b) the stand-alone deployment modes of NB-IoT.

2. **Formula used for calculation of the UL resource occupancy by the NPRACH**

$$T_{occ,i} = \frac{T_{preamb,i} * 100}{T_{period,i}} \quad \text{(F.1)}$$

$$F_{occ,i} = \frac{N_{preamb,i} * 100}{48} \quad \text{(F.2)}$$

$$UL_{occ,i} = T_{occ,i} * F_{occ,i} \quad \text{(F.3)}$$

$$UL_{occ} = \sum_{i=1}^{3} UL_{occ,i} \quad \text{(F.4)}$$

with

- $T_{occ,i}$ the time resource occupancy (%) of the NPRACH resources of coverage level **i**
- $F_{occ,i}$ the frequency resource occupancy (%) of the NPRACH resource of coverage level **i**
- $UL_{occ,i}$ the UL resource occupancy (%) of the NPRACH resources of coverage level **i**
- UL_{occ} the cumulative UL resource occupancy of the NPRACH resources of all coverage levels
- $T_{preamb,i}$ the time duration of the preamble for coverage level **i**
- $T_{period,i}$ the time duration of the NPRACH period for coverage level **i**
- $N_{preamb,i}$ the number of allocated preambles for coverage level **i**

References

[1] Ericsson AB, “Cellular networks for Massive IoT,” January 2016. [Online]. Available: https://www.ericsson.com/res/docs/whitepapers/wp_iot.pdf. [Accessed 4 April 2017].

[2] SigFox, “Sigfox - The Global Communications service provider for the Internet of Things (IoT),” [Online]. Available: https://www.sigfox.com/en. [Accessed April 2017].

[3] LoRa Alliance, “LoRa Technology - LoRa Alliance,” [Online]. Available: https://www.lora-alliance.org/What-Is-LoRa/Technology. [Accessed April 2017].

[4] 3GPP, “The 3rd Generation Parternship Project,” [Online]. Available: http://www.3gpp.org/. [Accessed April 2017].

[5] 3GPP, “TR 45.820 Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)”, version 13.1.0, December 2015. [Online]. Available: http://www.3gpp.org/.

[6] Qualcomm Technologies, Inc., “Paving the Path to Narrowband 5G with LTE IoT,” July 2016. [Online]. Available: https://www.qualcomm.com/documents/paving-path-narrowband-5g-lte-iot. [Accessed April 2017].

[7] U.S Department of Energy, “Smart Grid: An Introduction,” 2004. [Online]. Available: https://energy.gov/oe/downloads/smart-grid-introduction-0. [Accessed April 2017].

[8] “IEEE Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS), End-Use Applications, and Loads," in *IEEE Std 2030-2011*, vol., no., pp.1-126, Sept. 10, 2011.

[9] Open Smart Grid SG-Network task force, “Shared Documents - SG Communications (NIST PAP's),” May 2012. [Online]. Available: http://osgug.ucaiug.org/. [Accessed April 2017].

[10] M. Kuzlu, M. Pipattanasomporn, and S. Rahman, “Communication network requirements for major smart grid applications in HAN, NAN and WAN,” in *Computer Networks*, vol. 67, pp. 74–88, Jul. 2014.

[11] V. C. Gungor et al., "A Survey on Smart Grid Potential Applications and Communication Requirements," in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28-42, Feb. 2013

[12] Y. Jiang et al., "Outage Management of Distribution Systems Incorporating Information from Smart Meters," in *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 4144-4154, Sept. 2016.

[13] F. C. L. Trindade, W. Freitas and J. C. M. Vieira, "Fault Location in Distribution Systems Based on Smart Feeder Meters," in *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 251-260, Feb. 2014.

[14] R. Ratasuk et al., "NB-IoT system for M2M communication," *2016 IEEE Wireless Communications and Networking Conference*, Doha, 2016, pp. 1-5.

[15] A. Laya, L. Alonso, and J. Zarate, “Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives,” in *IEEE Communication Surveys & Tutorial*s, vol. 16, no. 1, pp. 4–16, 2014.

[16] SUNSEED consortium, “SUNSEED Deliverables,” 31 March 2016. [Online]. Available: http://sunseed-fp7.eu/wp-content/uploads/2015/04/SUNSEED-WP2-D212-V10-Final.pdf. [Accessed April 2017].

[17] Huawei Technologies, “NB-IoT - Enabling New Business Opportunites,” November 2016. [Online]. Available: https://gsacom.com/paper/nb-iot-enabling-new-opportunities/. [Accessed April 2017].

[18] International Telecommunication Union - Radiocommunications (ITU-R), “IMT Vision - Framework and overall objectives of the future deployment of IMT for 2020 and beyond,” 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf. [Accessed April 2017].

[19] Y. Shi, W. Xi and H. Zhou, “Efficient paging message design based on a binary tree in NB-IoT system,” *2016 IEEE Conference on Standards for Communications and Networking (CSCN)*, Berlin, 2016, pp. 1-6.

[20] Y. Wang and Z. Wu, “A coexistence analysis method to apply ACLR and ACS between NB-IoT and LTE for the stand-alone case,” *2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, Harbin, 2016, pp. 375-379.

[21] V. Petrov et al., “Vehicle-Based Relay Assistance for Opportunistic Crowdsensing over Narrowband IoT (NB-IoT),” in *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1-1, 2017.

[22] S.-M. Oh and J. Shin, “An Efficient Small Data Transmission Scheme in the 3GPP NB-IoT System,” in *IEEE Communication Letters*, vol. 7798, no. c, pp. 1-1, 2016.

[23] X. Lin, A. Adhikary and Y. E. Wang, “Random Access Preamble Design and Detection for 3GPP Narrowband IoT System,” in *IEEE Wireless Communication Letters*, vol. 5, no. 6, pp. 1-4, 2016.

[24] H. Kroll et al., “Maximum-likelihood detection for energy efficient timing acquisition in NB-IoT,” *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1-5, 2017.

[25] M. Lauridsen et al., “Coverage and Capacity Analysis of LTE-M and NB-IoT in a Rural Area,” *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Montreal, QC, 2016, pp. 1-5.

[26] C. Yu et al., “Uplink Scheduling and Link Adaptation for Narrowband Internet of Things Systems,” in *IEEE Access*, vol. 3536, no. c, pp. 1-1, 2017.

[27] S. Persia and L. Rea, "Next generation M2M Cellular Networks: LTE-MTC and NB-IoT capacity analysis for Smart Grids applications," *2016 AEIT International Annual Conference (AEIT)*, Capri, 2016, pp. 1-6.

[28] J.W. Shin et al., “Control Channel Load Balancing in Narrow Band Cellular IoT Systems Supporting Coverage Class”, *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, Bangkok, Thailand, 2016, pp. 343-348.

[29] SUNSEED EU, “D3.2.2 Design specification of communication solution for the smart grid support,” version 1.0, Deliverable report, pp. 16 -21, SUNSEED project, 2017.

[30] K. Sandrasegaran, H.A.M. Ramli and R. Basukala, “Delay-Prioritized Scheduling (DPS) for real time traffic in 3GPP LTE system”, *IEEE Wireless Communications and Networking Conference, WCNC*, 1–6. (2010).

[31] A.K.F. Khattab and K.M.F. Elsayed, “Channel-quality Dependent Earliest Deadline Due Fair Scheduling Schemes for Wireless Multimedia Networks”, 7th International Symposium on Modelling Analysis and Simulation of Wireless and Mobile Systems (MSWiM), Venice, Italy, Oct. 2004.

[32] I. M. Delgado-Luque et al., "Evaluation of latency-aware scheduling techniques for M2M traffic over LTE," *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, 2012, pp. 989-993.

[33] N. Afrin, J. Brown and J. Y. Khan, "A delay sensitive LTE uplink packet scheduler for M2M traffic," *2013 IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, 2013, pp. 941-946.

[34] W. Meng, R. Ma and H. H. Chen, "Smart grid neighborhood area networks: a survey," in *IEEE Network*, vol. 28, no. 1, pp. 24-32, January-February 2014.

[35] Department of Energy(DoE) USA, “Communication Requirements of Smart Grid Technologies”, Report, October 5, 2010. [Online]. Available : https://energy.gov/sites/prod/files/gcprod/documents/Smart_Grid_Communications_Requirements_Report_10-05-2010.pdf. [Accessed April 2017].

[36] C. Kalalas, L. Thrybom, and J. Alonso-Zarate, “Cellular Communications for Smart Grid Neighbourhood Area Networks: A Survey,” in *IEEE Access*, vol. 4, pp. 1469-1493, 2016.

[37] Y. Kabalci, “A survey on smart metering and smart grid communication,” in *Renewable and Sustainable Energy Reviews*, vol. 57, no. 302-318, 2016.

[38] J. J. Nielsen et al., “What can wireless cellular technologies do about the upcoming smart metering traffic?”, in *IEEE Communications Magazine*, vol. 53, no. 9, pp. 41-47, 2015.

[39] H. Quang-Dung, G. Yue, and L.N. Tho, “Challenges and research opportunities in wireless communication networks for smart grid,” in *IEEE Wireless Communications*, vol. 20, no. 3, pp. 89-95, 2013.

[40] 3GPP, “TS 36.300 Radio access network: Overall description”, version 13.4.0, June 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[41] Y. Wang *et al.*, "A Primer on 3GPP Narrowband Internet of Things", in *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117-123, 2017.

[42] 3GPP, "TS 36.101 UE radio transmission and reception", version 13.4.0, September 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[43] 3GPP, "TS 36.201 Physical layer general description", version 13.2.0, June 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[44] 3GPP, "TS 36.211 Physical channels and modulation", version 13.3.0, November 2016. [Online]. Available: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=50673,

[45] Analysys Mason, "Successful deployment of NB-IoT will depend on spectrum choices", January 2017. [Online]. Available: http://www.analysysmason.com/About-Us/News/Newsletter/successful-deployment-of-nb-iot-jan17/ . [Accessed: July 2017].

[46] J. Schlienz, and D. Raddino, "Narrowband Internet of Things whitepaper", Rohde & Schwarz, August 2016. [Online]. Available : https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_application/application_notes/1ma266/1MA266_0e_NB_IoT.pdf. [Accessed April 2017].

[47] 3GPP, "TS 36.213 Physical layer procedures", version 13.3.0, November 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[48] 3GPP, "TS 36.212 Multiplexing and channel coding", version 13.3.0, October 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[49] 3GPP, "TS 36.331 Radio Resource Control (RRC); Protocol specification", version 13.3.0, October 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[50] 3GPP, "TS 36.321 Medium Access Control (MAC) protocol specification", version 13.3.0, October 2016. [Online]. Available: http://www.3gpp.org/DynaReport/WiVsSpec--690063.htm.

[51] Ericsson, "R1 - 161835: NB-IoT – NPRACH configurations," Written contribution, 3GPP TSG RAN1 NB-IoT Ad-Hoc meeting, Sophia Antipolis, France, March 2016.

[52] R. G. Cheng et al., "RACH Collision Probability for Machine-Type Communications," *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, Yokohama, 2012, pp. 1-5.

[53] Matrix Laboratory (MATLAB) Release 2015a, *The MathWorks Inc.*, Massachusetts, USA.

[54] 3GPP, ''TR 25.942 Radio Frequency (RF) system scenarios (Release 13),'' January 2016, [Online]. Available: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=48837.

[55] 3GPP, ''TR 25.816 UMTS 900 MHz Work Item Technical Report (Release 7),'' December 2012, [Online]. Available:https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1249.

[56] M. Nijhuis et al., "Demand Response: Social Welfare Maximization in an Unbundled Energy Market Case Study for the Low-Voltage Networks of a Distribution Network Operator in The Netherlands"", *IEEE Transactions on Industry Applications*, vol. 53, no. 1, pp. 32–38, 2017.

[57] M. Nijhuis and J.F.G. Cobben, "Effects of power quality limits on LV-network design", *17th International Conference on Harmonics and Quality of Power (ICHQP)*, pp. 83–88, 2016.

[58] StatLine, Centraal Bureau voor de Statistiek (CBS), The Netherlands. Available: https://statline.cbs.nl/Statweb. [Accessed June 2017].

[59] R. Litjens et al., "Assessment of the energy efficiency enhancement of future mobile networks," *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, 2014, pp. 2336-2341.

[60] F. Gunnarsson et al., "Downtilted Base Station Antennas - A Simulation Model Proposal and Impact on HSPA and LTE Performance," *2008 IEEE 68th Vehicular Technology Conference*, Calgary, BC, 2008, pp. 1-5.

[61] ZTE, "R1 - 160480: Consideration on uplink data transmission for NB-IoT," Written contribution, 3GPP TSG RAN1 NB-IoT meeting #84, St Julian's, Malta, 15-19 February 2016. [Online].

Available:https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=683004. [Accessed March 2017].

[62] ZTE, "R1 - 161860: Further considerations on NB-PDSCH design for NB-IoT," Written contribution, 3GPP TSG RAN1 NB-IoT Ad-Hoc meeting #2, Sophia Antipolis, France, 22-24 March 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_AH/LTE_NB-IoT_1603/Docs/. [Accessed March 2017].

[63] L. E. Miller and J. S. Lee, "BER expressions for differentially detected pi/4 DQPSK modulation," in *IEEE Transactions on Communications*, vol. 46, no. 1, pp. 71–81, 1998.

[64] 3GPP, "TR 37.868 Study on RAN Improvements for Machine-type Communications (Release 11)," September 2011. [Online]. Available : https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2630.

[65] R.H. Arpaci-Dusseau and A.C. Arpaci-Dusseau, " Operating Systems : Three Easy Pieces ", Arpaci-Dusseau Books, Version 0.90, March 2015.

[66] M. Harchol-Balter et al.," Size-based scheduling to improve web performance", ACM Transactions on Computer Systems, vol. 21, no.2, pp. 207-233, 2003.

[67] S.H. Khan, "Capacity study for LTE release 13 eMTC with coverage enhanced RACH", Master thesis, submitted to the Eindhoven University of Technology, The Netherlands, August 2016. [Online]. Available: https://pure.tue.nl/ws/files/46944860/855330-1.pdf. [Accessed April 2017].

[68] M. Amirijoo, F. Gunnarsson, and F. Andren, "3GPP LTE random access channel self-optimization," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 2784–2793, July 2014.

[69] S. Cherkaoui et al., "LTE-A random access channel capacity evaluation for M2M communications," *2016 Wireless Days (WD)*, Toulouse, 2016, pp. 1-6.

[70] F. Alsewaidi, A. Doufexi and D. Kaleshi, "Enhancing Radio Access Network Performance over LTE-A for Machine-to-Machine Communications under Massive Access", in *Mobile Information Systems*, vol. 2016, Article ID 5187303, 16 pages, 2016.

[71] Intel Corporation, "R1-161893: Remaining details of NB-IoT multi-carrier operation", Written contribution, 3GPP TSG RAN WG1 NB-IoT Ad-Hoc Meeting # 2, Sophia Antipolis, France, March 22-24 2016. [Online].
Available:http://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=692206.
[Accessed January 2017].

[72] Huawei and HiSilicon, "R1-156923: Multiple NB-IoT carrier deployment", Written contribution, 3GPP TSG RAN WG1 Meeting #83, Anaheim, USA, November 15-22 2015. [Online].
Available: https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=667927 .
[Accessed January 2017].

[73] A. Lo et al, "Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine Communications", in Proc. of *27th Meeting of Wireless World Research Forum (WWRF),* Oct. 2011.

[74] J.P. Cheng, C. Han Lee, and T.-M. Lin, "Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks," *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, TX, 2011, pp. 368-372.

Abbreviations

ACK	Acknowledgement
AMI	Advanced Metering Infrastructure
AWGN	Additive White Gaussian Noise (channel)
BLER	Block Error Rate
BPL	Building Penetration Loss
BPSK	Binary Phase Shift Keying
BSR	Buffer Status Report
COST	European Cooperation in Science and Technology
CP	Cyclic Prefix
CQI	Channel Quality Indicator
CRC	Cyclic Redundancy Check
DA	Distribution Automation
DCI	Downlink Control Information
DCS	Distribution Customer Storage
DL	Downlink
DLC	Direct Load Control
DMRS	Demodulation Reference Signal
DR	Demand Response
DSO	Distribution Service Operator
DV	Data Volume
EDDF	Earliest Due Date First
eNodeB	Evolved Node B
ET	Electric Transportation
5G	Fifth Generation
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HAN	Home Area Network
HD-FDD	Half Duplex – Frequency Division Duplexing
HEMS	Home Energy Management System
HH	Household
HOL	Head-of-Line
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IoT	Internet of Things
ISD	Inter-Site Distance
KPI	Key Performance Indicator
LoRaWAN	Long Range Wide Area Network
LPWA	Low Power Wireless Access
LTE	Long Term Evolution
LTE-A	Long Term Evolution - Advanced
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MME	Mobility Management Entity
MNO	Mobile Network Operator
MSG	Message
NACK	Negative Acknowledgement
NAN	Neighbourhood Area Network
NAS	Non-Access Stratum
NB-IoT	Narrowband Internet-of-Things
NCCE	Narrowband Control Channel Element
NLOS	Non-Line-of-Sight
NPDCCH	Narrowband Physical Downlink Control Channel
NPDSCH	Narrowband Physical Downlink Shared Channel

NPRACH	Narrowband Physical Random Access Channel
NPUSCH	Narrowband Physical Uplink Shared Channel
NPSS	Narrowband Primary Synchronisation Signal
NRS	Narrowband Reference Signal
NSSS	Narrowband Secondary Synchronisation Signal
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OpenSG	Open Smart Grid
ORM	Outage Restoration and Management
PHY	Physical
PLC	Power Line Communication
PMU	Phasor Measurement Unit
PRB	Physical Resource Block
PSD	Power Spectral Density
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAO	Random Access Opportunity
RAR	Random Access Response
RAT	Radio Access Technology
RE	Resource Element
RF	Radio Frequency
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RTP	Real-Time Pricing
RU	Resource Unit
S-GW	Serving Gateway
SC-FDMA	Single Carrier – Frequency Division Multiple Access
SF	Subframe
SINR	Signal to Noise plus Interference Ratio
SO	Scheduling Opportunity
SPTF	Shortest Processing Time First
SRS	Sounding Reference Signals
3GPP	The 3rd Generation Partnership Project
TB	Transport Block
TBS	Transport Block Size
TOU	Time-of-Use
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
WAMS	Wide Area Management and Supervision
WAN	Wide Area Network
WiMAX	Worldwide Interoperability for Microwave Access