Wireless network dimensioning and provisioning for ultra-reliable communication: Modeling and analysis

André Vinícius Gomes Santos Gonçalves

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Engineering

Luiz A. DaSilva, Chair

Jacek Kibiłda, Co-chair

Harpreet S. Dhillon

Jeffrey H. Reed

Shaddi Hasan

November 6, 2023

Arlington, Virginia

Keywords: Network reliability, Ultra-reliable communication, URLLC, HRLLC, spectrum dimensioning, meta distributions, extreme-value theory

Wireless network dimensioning and provisioning for ultra-reliable communication: Modeling and analysis

André Vinícius Gomes Santos Gonçalves

(ABSTRACT)

A key distinction between today's and tomorrow's wireless networks is the appetite for reliability to enable emerging mission-critical services such as ultra-reliable low-latency communication (URLLC) and hyper-reliable low-latency communication (HRLLC), the staple mission-critical services in IMT-2020 (5G) and IMT-2023 (6G), for which reliable and resilient communication is a must. However, achieving ultra-reliable communication is challenging because of these services' stringent reliability and latency requirements and the stochastic nature of wireless networks. A natural way of increasing reliability and reducing latency is to provision additional network resources to compensate for uncertainty in wireless networks caused by fading, interference, mobility, and time-varying network load, among others. Thus, an important step to enable mission-critical services is to identify and quantify what it takes to support ultra-reliable communication in mobile networks – a process often referred to as dimensioning. This dissertation focuses on resource dimensioning, notably spectrum, for ultra-reliable wireless communication. This dissertation proposes a set of methods for spectrum dimensioning based on concepts from risk analysis, extreme value theory, and meta distributions. These methods reveal that each "nine" in reliability (e.g., five-nines in 99.999%) roughly translates into an order of magnitude increase in the required bandwidth. In ultra-reliability regimes, the required bandwidth can be in the order of tens of gigahertz, far beyond what is typically available in today's networks, making it

challenging to provision resources for ultra-reliable communication. Accordingly, this dissertation also investigates alternative approaches to provide resources to enable ultra-reliable communication services in mobile networks. Particularly, this dissertation considers multi-operator network sharing and multi-connectivity as alternatives to make additional network resources available to enhance network reliability and proposes multi-operator connectivity sharing, which combines multi-operator network sharing with multi-connectivity. Our studies, based on simulations, real-world data analysis, and mathematical models, suggest that multi-operator connectivity sharing – in which mobiles multi-connect to base stations of operators in a sharing arrangement – can reduce the required bandwidth significantly because underlying operators tend to exhibit characteristics attractive to reliability, such as complementary coverage during periods of impaired connectivity, facilitating the support for ultra-reliable communication in future mobile networks.

Wireless network dimensioning and provisioning for ultra-reliable communication: Modeling and analysis

André Vinícius Gomes Santos Gonçalves

(GENERAL AUDIENCE ABSTRACT)

A key distinction between today's and tomorrow's wireless networks is the appetite for reliability to enable emerging mission-critical services in 5G and 6G, for which ultra-reliable communication is a must. However, achieving ultra-reliable communication is challenging because of these services' stringent reliability and latency requirements and the stochastic nature of wireless networks. Reliability often comes at the cost of additional network resources to compensate for uncertainty in wireless networks. Thus, an important step to enable ultra-reliable communication is to identify and quantify what it takes to support mission-critical services in mobile networks – a process often denoted as dimensioning. This dissertation focuses on spectrum dimensioning and proposes a set of methods to identify suitable spectrum bands and required bandwidth for ultra-reliable communication. These methods reveal that the spectrum needs for ultra-reliable communication can be beyond what is typically available in today's networks, making it challenging to provide adequate resources to support ultra-reliable communication services in mobile networks. Alternatively, we propose multi-operator connectivity sharing: mobiles simultaneously connect to multiple base stations of different operators. Our studies suggest that multi-operator connectivity sharing can reduce the spectrum needs in ultra-reliability regimes significantly, being an attractive alternative to enable ultra-reliable communication in future mobile networks.

Dedication

To my family and in memory of my grandfather, Ademar Gomes da Silva.

Acknowledgments

First and foremost, I want to express my deepest gratitude to my advisors, Dr. Luiz DaSilva and Dr. Jacek Kibilda, for all their support and guidance. The Ph.D. journey can sometimes be bitter. Fortunately, I feel extremely privileged to belong to the group of Ph.D. graduates who have genuinely enjoyed their journey and found in their advisors a role model for what they can only aspire to be as academics and mentors. Luiz and Jacek, you have set the bar high! I also want to thank them for their support during my transfer from Trinity College Dublin (TCD), Ireland, to Virginia Tech in 2021, helping me navigate uncertain and adverse COVID-19 times. I sincerely thank my committee members for their valuable input. In particular, I want to thank Dr. Harpreet, whose feedback and engagement during my Preliminary Exam and teaching in ECE 5605 inspired my latest works on meta distributions and unlocked a new set of skills I would have hardly considered otherwise. His enthusiasm and clarity in teaching inspire me for what I want to be as a professor. Furthermore, I want to thank my friends and colleagues who have supported me in my Ph.D. journey. To start with, I thank my dear friends and colleagues at TCD, including Ayush (whose friendship and travel companion I have been fortunate to carry to the US), Abel (who shares the same convoluted sense of direction as me, resulting in many incredible, longer-than-expected hikes because of wrong turns), Fadhil, Erika, Merim, Nima, and Jean-Baptiste. I also thank Dr. Nicola Marchetti for assisting me during my transfer to Virginia Tech, temporarily hosting me as one of his students at TCD. Similarly, I thank my dear friends and colleagues at Virginia Tech, especially Vikas (for endless conversations, bike rides, and hikes), Biplav (for his help even before I officially joined Virginia Tech), Jefferson (for his commitment to our "doctors of fitness" gym project), João, Prateek, Adytia, Adrian, and Nazmul.

Contents

List of Figures

List of Tables

List of Abbreviations

BS Base station

CDF Cumulative density function

eMBB Enhanced MBB

EVT Extreme value theory

FDMA Frequency division multiple access

HRLLC Hyper-reliability low-latency communication

KDE Kernel density estimation

KPI Key performance indicator

MBB Mobile broadband

MNO Mobile network operator

PDF Probability density function

PPP Poisson point process

RSRP Reference signal received power

RSRQ Reference signal received quality

SINR Signal-to-interference-plus-noise ratio

SIR Signal-to-interference-ratio

SNR Signal-to-noise ratio

URLLC Ultra-reliable low-latency communication

Chapter 1

Introduction

1.1 Motivation and overview

Perhaps the most challenging goal for the next generation of mobile networks is to achieve ultra-reliable communication to enable emerging mission-critical services in factory automation, augmented reality, cloud gaming, smart transportation, and yet-to-come applications, for which ultra-reliable and resilient wireless communications is a must [1, 2, 3]. Ultra-reliable communication is challenging because of the stringent reliability and latency requirements of ultra-reliable communication services as well as the stochastic nature of wireless networks, subject to spatiotemporal randomness caused by phenomena such as user mobility, fading, interference, and varying network load. As an example, in ultra-reliable low-latency communication (URLLC) and hyper-reliable low-latency communication (HRLLC), the staple ultra-reliable communication services in IMT-2020 [4] and (recently released drafts for) IMT-2030 [5], reliability is a function of "nines," typically 5-7 (e.g., five-nines is 99.999%), and latency less than a few milliseconds, making ultra-reliable communication services sensitive to even rare network conditions that may only occur with probability $\leq 0.001\%$ when reliability is $\geq 99.999\%$.

A natural way to increase reliability, reduce latency, and compensate for uncertainty in wireless communications is to provision additional network resources, be it in the form of

additional antennas (e.g., multi-antenna communication in [6]), spectrum (e.g., wider channels in [6] and channel hopping in [7, §III.B.3]), base stations (e.g., network densification in [2]), or connections (e.g., multi-connectivity in [8, 9]). Similarly, several works focus on the prioritization of ultra-reliable communication services, which is also a form of making additional network resources available, as more resources are reserved for ultra-reliable communication services compared to other types of services (e.g., scheduling prioritization of URLLC traffic over eMBB traffic in [10]). This corroborates the anecdotal evidence that, for whatever resource it is, we need more of it to support high reliability and low latency [11]. However, blindly adding resources to the network is costly, thus it is important to ask *what level of resources is minimally required for mobile networks to support the desired service requirements*, a process often referred to as dimensioning.

The challenge to answer that question begins with defining the relationship between ultra-reliable communication service requirements and resource requirements. Unlike eMBB, for which IMT-2020 specifies pre-defined target average, peak, and 5th percentile data rate and spectral efficiency goals as well as a set of KPI formulas that map them onto bandwidth (a network resource), the service requirements for URLLC and HRLLC are uniquely defined in terms of higher-level goals, namely latency and reliability, without straightforward methods to map them onto network resources [4]. The challenge continues with the provision of resources. Intuition suggests that the level of resources needed to compensate for stochastic wireless network phenomena and guarantee tight deadlines at the confidence level $\geq 99.999\%$ may be substantial. In fact, as we shall discuss in detail in this dissertation, meeting ultra-reliable communication service goals throughout the network may require bandwidth in the order of gigahertz or massive densification (e.g., ≈ 100 base stations per km^2), beyond what is typically available (or even practical) in today's mobile networks.

The main objective of this dissertation is to address such challenges and provide insights on

network dimensioning and provisioning for network-wide ultra-reliable communication. In the next section, we will look closer at network dimensioning for ultra-reliable communication and potential approaches to facilitate the provision of resources to meet ultra-reliability goals. Next, Section 1.3 highlights this dissertation's main contributions and lists the related publications achieved as part of my Ph.D. Section 1.4 outlines the remainder chapters of this dissertation.

1.2 Background and selected prior works

1.2.1 Network dimensioning for ultra-reliable communication

Network dimensioning for ultra-reliable communication has been only addressed by a handful of prior works. Yet, they mostly focus on micro dimensioning, as in [12], where the authors estimated the bandwidth requirements to meet ultra-reliable communication service requirements for given signal-to-noise ratio values, and [6], where the authors estimated the bandwidth and antenna configuration requirements for ultra-reliable communication in small-scale industrial networks (e.g., a network with a single base station covering an area much smaller than 1 km^2). While relevant for scheduling, rate control, and coding/decoding mechanisms as well as factory applications, these works provide limited insights into the level of resources required to support ultra-reliable communication network-wide, where channel conditions typically vary widely.

The stringent service requirements for ultra-reliable communication and the stochastic nature of wireless networks make network dimensioning particularly challenging. First, as opposed to eMBB, there are no straightforward ways to map the relationship between ultra-reliable communication service requirements and network resource requirements [4]. Second, in ultra-

reliability regimes, reliability is often 99.999% or higher [4], meaning that ultra-reliable communication services are sensitive to even rare network conditions that may only occur with probability 0.001% or lower, which can be hard to capture or assess. In response, [13] surveyed frameworks to capture and model network asymptotics that pertain to ultra-reliable communication service requirements. In our research, we have explored the following frameworks:

- Risk analysis is a framework to model and analyze low-probability events, being widely used in financial engineering to assess low-probability losses. Popular measures of risk include quantiles (aka value-at-risk) and superquantiles (aka conditional value-at-risk) [14, 15]. Mathematically, quantiles are simply the inverse of a distribution function, whereas superquantiles are the expectation over a given threshold, i.e., $\mathbb{E}[X|X \geq x]$, where the threshold x is generally a quantile, providing information about the expected worst-case scenarios.

- Extreme value theory is a branch of statistics that focuses on the extreme tails of a distribution [16]. Notably, the Pickands-Balkema-De Haan theorem postulates that the cumulative distribution function of exceedances over a sufficiently high threshold tends to the generalized Pareto distribution for a large class of distribution functions, being generally used to characterize extreme low-probability events. In the context of ultra-reliable communication, this theorem can be used to characterize extremely low-probability network conditions, as conditions that may only occur $\leq 0.001\%$ of the time.

- Meta distributions are defined as the distribution function of a conditional distribution [17, 18], e.g., $\Pr\left(\Pr\left(f(X,Y) \geq \theta|Y\right) \geq x\right)$ for $x \in [0,1]$, and provide finer-grained information about network asymptotics under strict reliability constraints than standard

success probability measures, e.g., $\Pr(f(X,Y) \geq \theta)$. In fact, standard success probability measures can be obtained by taking the expectation over meta distributions, i.e., $\Pr(f(X,Y) \geq \theta) = \mathbb{E}\left[\Pr\left(f(X,Y) \geq \theta | Y\right)\right]$.

These frameworks have been used in the context of ultra-reliable communication in several ways, as in [19], where superquantiles are applied to the problem of resource scheduling for eMBB and URLLC services, [20, 21], where extreme value theory is used to (a) approximate low-probability queue delays in a transmit power minimization problem subject to reliability and latency constraints and (b) model the tail statistics of the received signal power for in-vehicle wireless communications, and [22], where meta distributions were used to obtain the asymptotics of the SIR in Poisson networks with Rayleigh fading. Similarly, these frameworks can potentially drive network dimensioning for ultra-reliable communication and will, in fact, be explored in subsequent chapters of this dissertation.

1.2.2 Facilitators of ultra-reliable communication

Multi-connectivity. Redundancy is a common and natural approach to improve reliability. Perhaps the most common form of redundancy in networking is timely retransmissions, such as in the transmission control protocol (TCP), where unacknowledged packets are retransmitted after a timeout [23]. This approach, however, does not cope well with the stringent latency requirements of ultra-reliable communication services, where deadlines are typically a few milliseconds or less [4, 5]. Alternatively, one could reduce the timeouts for retransmissions or even enable proactive retransmissions without a timeout, where duplicates are transmitted in consecutive time slots. Yet, timely retransmissions can be ineffective because consecutive (or close in time) slots are often correlated, meaning that failures are likely to persist for several time slots, resulting in no significant reliability gains [24]. In this re-

gard, diversity is key to redundancy: there are no significant gains in reliability if redundant resources are strongly and positively correlated.

Redundancy can come in other forms, such as multi-connectivity (aka interface diversity), which increases network reliability by providing parallel, redundant paths for communication. In multi-connectivity, diversity is often achieved by using independent connections, such as connections to base stations far apart (i.e., spatial diversity) or operating in different frequency bands (i.e., spectral diversity). Multi-connectivity has been studied in many papers as a key enabler of ultra-reliable communication (consider [25] for a comprehensive list). Notably, [9] studied multi-connectivity at the network/transport layers, where information is split into packets and transmitted simultaneously over several radio interfaces, e.g., 3G, 4G, and WiFi. The authors explored several encoding/splitting techniques. In practical scenarios, packet duplication, where packets are fully replicated over all interfaces, significantly reduces latency while increasing network reliability, outperforming more sophisticated packet-splitting strategies. [8], in turn, studied multi-connectivity at the physical (PHY) layer, where signals of different base stations are combined according to techniques such as selection combining and maximal-ratio combining. In both cases, multi-connectivity is shown to decrease the outage probability, being beneficial to ultra-reliable communication.

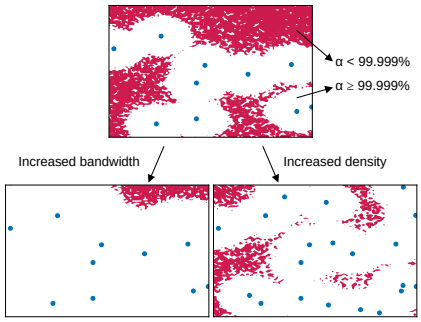
Network sharing. The main benefits of network sharing for ultra-reliable communication are twofold. First, network sharing is a cost-effective way of increasing network resources without outright investments [26, 27], which is beneficial for ultra-reliable communication because, as discussed in the previous section, reliability often comes at the cost of additional network resources to compensate for uncertainty in wireless networks. Second, multi-operator network sharing (as infrastructure sharing in [28] or akin to mobile virtual network operators such as Google in the US, which operates on top of a multi-operator network) generally exhibits increased diversity because underlying operators in the sharing agreement

often operate in their own licensed bands (i.e., spectrum diversity) and infrastructure (i.e., spatial diversity). Furthermore, [29] provides empirical evidence that mobile operators exhibit complementary spatiotemporal characteristics, such as in network load, reducing the number of hot spots in multi-operator network sharing (i.e., locations with high enough load that can throttle resources and jeopardize communication).

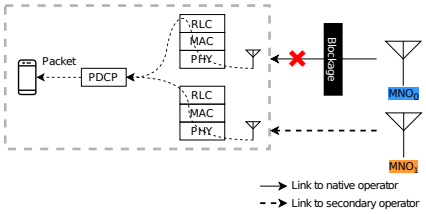
1.3 Contributions

This dissertation addresses the challenge of wireless network dimensioning and provisioning for network-wide ultra-reliable communication. The core contributions of this dissertation are summarized below.

New methods for network dimensioning. We develop a set of empirical, algorithmic, and analytical methodologies based on risk analysis, extreme value theory, and meta distributions, respectively, to map ultra-reliable communication service requirements onto network resource requirements such as spectrum and network density. Using our methods, we can, for instance, quantify the amount of spectrum/bandwidth it takes to support ultra-reliable communication services at the α-reliability level in $\eta\%$ of the network area (i.e., the percentage corresponding to the service area in white in Figure 1.1a). Our findings reveal an interesting relationship between service and spectrum requirements, where each "nine" in reliability can roughly translate into one order of magnitude increase in bandwidth. As a practical contribution, our methodologies for spectrum dimensioning are expected to contribute to the Next G Alliance – an industry-led initiative to advance a North American vision for the next generations of mobile networks (nextgalliance.org) – guidelines for assessing spectrum needs for 6G.



(a) Blue markers represent base stations. White regions support ultra-reliable communication.



(b) Multi-operator connectivity sharing in a dual-operator network.

Figure 1.1: Illustration of the main areas of contribution in this dissertation, to be further explored in later chapters.

Multi-operator connectivity sharing. In ultra-reliability regimes, the required bandwidth can be in the order of gigahertz, beyond what is typically available in today's networks. Inspired by multi-connectivity and network sharing, we propose multi-operator connectivity sharing for ultra-reliable communication to facilitate the provisioning of resources for ultra-reliable communication. In multi-operator network sharing, mobiles multi-connect to base stations of operators in a sharing arrangement as illustrated in Figure 1.1b. Using simulation and analytical models, we show that multi-operator connectivity sharing can significantly reduce the demand for spectrum without outright investments in new infrastructure such as new base stations. Based on a real-world study, we demonstrate that multi-operator connectivity sharing is attractive because underlying operators tend to exhibit complementary characteristics, such as complementary coverage during periods of impaired connectivity. However, multi-operator connectivity sharing, as multi-connectivity in general, incurs an increased network load because of the additional connections, which may throttle the gains caused by complementary coverage. We show that a simple but effective strategy of only allowing users with weak signal strength to multi-connect can significantly increase reliability at a minor penalty to the overall performance.

1.3.1 List of publications

This dissertation led to the journal (label J) and conference (label C) papers listed below.

J1. **A. Gomes**, J. Kibiłda and L. A. DaSilva, "Assessing the spectrum needs for network-wide ultra-reliable communication with meta distributions," in IEEE Communications Letters, vol. 27, no. 8, pp. 2242-2246, Aug. 2023, DOI: 10.1109/LCOMM.2023.3282140.

J2. **A. Gomes**, J. Kibiłda, N. Marchetti and L. A. DaSilva, "Dimensioning spectrum to support ultra-reliable low-latency communication," in IEEE Communications Standards Magazine, vol. 7, no. 1, pp. 88-93, Mar. 2023, DOI: 10.1109/MCOMSTD.0004.2100107.

J3. **A. Gomes**, J. Kibiłda and L. A. DaSilva, "Capturing rare network conditions to dimension resources for ultra-reliable communication," in IEEE Communications Letters, vol. 26, no. 11, pp. 2789-2793, Nov. 2022, DOI: 10.1109/LCOMM.2022.3202841

J4. **A. Gomes**, J. Kibiłda, A. Farhang, R. Farrell and L. A. DaSilva, "Multi-operator connectivity sharing for reliable networks: A data-driven risk analysis," in IEEE Transactions on Network and Service Management, vol. 18, no. 3, pp. 2800-2811, Sep. 2021, DOI: 10.1109/TNSM.2021.3073841

C5. **A. Gomes**, J. Kibiłda, and L. A. DaSilva, "Analysis of spectrum needs for network-wide ultra-reliable communication with network sharing," in IEEE Globecom Workshops, 2023. (To appear)

C6. **A. Gomes**, J. Kibiłda, and L. A. DaSilva, "Network sharing for reliable networks: A data-driven study," in IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020, pp. 1-6, DOI: 10.1109/ICC40277.2020.9148763

The papers above map onto this dissertation's technical chapters as in Table 1.1.

Chapter	Papers
Chapter 2	J2
Chapter 3	J4 and C6
Chapter 4	J3
Chapter 5	J1 and C5

Table 1.1: Mapping chapters onto their respective reference papers.

1.4 Organization

The remainder of this dissertation is organized into four technical chapters and a conclusion chapter, summarized below.

Chapter 2 uses 3GPP urban network models to draw the relationship between reliability requirements and two network assets, spectrum and network density. The required bandwidth in ultra-reliability regimes can be in the order of gigahertz. In our model, bandwidth can be traded for network density but comes at the cost of massive densification. Alternatively, we show that multi-operator connectivity sharing can reduce the bandwidth requirement by several orders of magnitude without massive network densification.

Chapter 3 advances the analysis of multi-operator connectivity sharing. We make use of real-world datasets of three mobile operators in Dublin, Ireland to assess the pros and cons of multi-operator connectivity sharing. Our analysis, based on superquantiles, indicates that multi-operator connectivity sharing increases reliability because underlying operators in the sharing arrangement tend to exhibit complementary characteristics, such as complementary coverage, which can reduce periods of connectivity shortfalls in multi-operator connectivity sharing. On the other hand, multi-operator connectivity sharing, as multi-connectivity in general, increases the network load because of the additional connections, which, in turn, can

throttle complementary coverage gains and affect performance. Our findings indicate that a simple but effective strategy of only allowing mobiles with weak signal strength to multi-connect leads to a significant increase in reliability at a small loss in overall performance.

Chapter 4 addresses the practical challenge of empirically assessing ultra-reliability regimes for network dimensioning. As an example, a rule of thumb for confident estimation is to have at least 100 likely instances of an event, implying that capturing network conditions that may only occur with probability 0.001% requires the collection of $\geq 10^7$ data samples (or simulation cycles) to dimension network resources when reliability is 99.999%, which can be impractical depending on the time it takes to collect each sample. To circumvent this challenge, we propose an extreme value theory-based algorithm for spectrum dimensioning for ultra-reliable communication. Our analysis shows that the proposed approach can reduce the demand for massive data by several orders of magnitude in ultra-reliability regimes, outperforming comparable state-of-the-art methods.

Chapter 5 advances spectrum dimensioning for network-wide ultra-reliable communication by proposing an analytical methodology to capture the relationship between spectrum needs and reliability target requirements. Our methodology is based on meta distributions and differentiates between two notions of reliability, namely link reliability and spatial reliability, corresponding to the success rate of meeting performance requirements of ultra-reliable communication services and the service area, in percentage, where ultra-reliable communication is supported in the network. Using the proposed methodology, we obtain closed-form expressions for the relationship between spectrum needs and reliability target requirements. Interestingly, our findings reveal a simple relationship, where each "nine" in reliability roughly maps onto an order of magnitude increase in bandwidth. It also reveals how multi-operator connectivity sharing can ease the amount of spectrum needed to meet network-wide ultra-reliability goals, indicating that even sharing arrangements with only two operators, when

coupled with multi-connectivity, can reduce the required bandwidth by several orders of magnitude.

Lastly, Chapter 6 summarizes the key findings of this dissertation and lists potential future research directions. It also presents additional details on research activities I have been involved in during my Ph.D. beyond what is presented in this dissertation.

Chapter 2

Dimensioning spectrum to support ultra-reliable low-latency communication

2.1 Introduction

As new communication services are identified, it is important to understand what level of network resources is needed to support their service-level requirements, a process often referred to as dimensioning. Particularly, spectrum is the lifeblood of wireless communications, and identifying and quantifying spectrum needs is key for future licensing, regulatory, and network deployment plans. In response, initiatives such as the Next G Alliance, an industry-led initiative to advance a North American vision for the next generations of mobile networks, are currently considering how to dimension the spectrum for new classes of services envisioned beyond 5G [30]. This, however, can be challenging for ultra-reliable communication services such as URLLC because of their stringent reliability and latency requirements. For instance, a typical URLLC requirement specified by the 3GPP is the transmission of 32 bytes within a 1 ms deadline with a 99.999% success probability, making URLLC sensitive to even rare events that occur with the probability of 0.001%.

This chapter addresses spectrum dimensioning for ultra-reliable communication services in large-scale mobile networks (i.e., outdoor and multi-cell network scenarios). To that end, we first map the ultra-reliable communication requirements onto bandwidth requirements and then assess the magnitude and type of spectrum needed to meet the service-level requirements on urban-macro network models developed by the 3GPP. As illustrated in Figure 2.1, making more bandwidth available compensates for cell edge effects and expands the reach of ultra-reliable communication. However, our results indicate that supporting ultra-reliable communication services at large-scale networks can demand bandwidth in the order of gigahertz, beyond what is typically available in today's networks. We study how network densification can expand the reach of ultra-reliable communication services without additional bandwidth, as illustrated in the right side of Figure 2.1. We then study network sharing and multi-connectivity as alternative ways to further ease the demand for bandwidth without outright deployment of additional base stations.

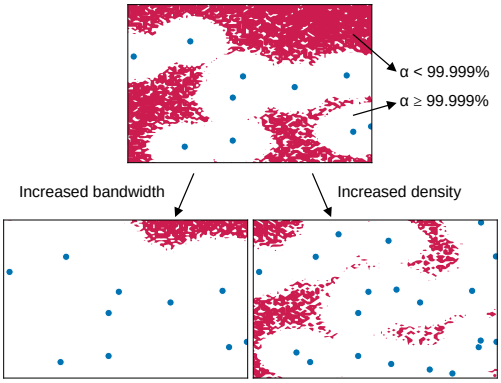


Figure 2.1: Blue markers represent base stations. Additional bandwidth or base stations expand the reach of ultra-reliable communication (white zones).

2.1.1 Related work

Spectrum dimensioning for ultra-reliable communication has only been considered in a handful of prior works, such as [12], which estimated the bandwidth to meet ultra-reliability goals given signal-to-noise ratio values, and [6], which estimated the bandwidth required to meet ultra-reliability goals in small-scale industrial networks (i.e., single-cell network scenarios

with area ≤ 1 km^2). While relevant for scheduling, rate control, and coding/decoding mechanisms as well as factory applications, these works provide limited insights regarding the resources needed to achieve ultra-reliable communication in large-scale mobile networks, where channel quality typically varies widely throughout the network. This chapter moves forward by considering the resource requirements for ultra-reliable communication in urban-macro network scenarios developed by the 3GPP.

2.2 Preliminaries

2.2.1 Expressing reliability

We adopt the ITU-R/3GPP definition of reliability for URLLC [4, 31]: the success probability α of transmitting δ bits of data within a user plane latency deadline τ, i.e., $\Pr(T \leq \tau) \geq \alpha$. We consider a typical mobile-to-base station link, without intermediate relays. The user plane latency T, i.e., the time it takes to transmit δ bits from/to the base station to/from the mobile in the downlink/uplink communication [31], is a sum of physical and data link layers' delays, which generally includes (a) a processing delay, incurred by signal processing; (b) a medium access delay, incurred by scheduling or contention for the medium; and (c) a transmission delay, i.e., $T = T_{\text{proc}} + T_{\text{ma}} + T_{\text{tx}}$. The processing delay is fairly constant and hardware/implementation related; the medium access delay can be deterministic (e.g., scheduling-based approaches) or random (e.g., contention-based approaches). In ultra-reliable communication, deterministic medium access is likely to be adopted, given the stringent reliability and latency requirements.

Hence, the transmission delay is the prime source of uncertainty given the probabilistic nature of wireless communication, prone to stochastic channel conditions caused by user mobility,

blockages, fading, and interference. We adopt a general formulation where the transmission delay relates to the Shannon-Hartley capacity and assume no protocol overhead. This way, we can estimate the *fundamental minimum amount of spectrum* needed to support ultra-reliable communication services. The transmission delay is expressed as:

$$T_\delta = \frac{\delta}{w \times \log_2(1+\Gamma)}, \tag{2.1}$$

where w and Γ are the bandwidth and SINR, respectively. We assume the SINR to be static during time τ, which is a reasonable assumption for small values of τ, such as in URLLC, where τ is often 1 ms.

With the focus on the transmission delay, the ultra-reliable communication goals can be expressed as:

$$\Pr\left(w \geq \frac{\delta}{\tau_{\text{tx}} \times \log_2(1+\Gamma)}\right) \geq \alpha, \tag{2.2}$$

where τ_{tx} is the fraction of the latency budget τ reserved for the transmission delay. Hence, the minimum bandwidth to satisfy the ultra-reliable communication service-level requirements is a function of the α-quantile Q_α of the inverse of the spectrum efficiency,

$$w = \frac{\delta}{\tau_{\text{tx}}} \times Q_\alpha\left(\frac{1}{\log_2(1+\Gamma)}\right). \tag{2.3}$$

2.2.2 The network model

We adopt the 3GPP urban-macro network model in TR 38.901 [32]. The model defines (a) antenna array configurations for base stations [32, §7.3]; (b) path loss [32, §7.4]; (c) fading [32, §7.4]; and (d) line-of-sight probabilities [32, §7.4] for different frequency bands ranging from 0.5 to 100 GHz. We consider three frequency bands centered at 700 MHz, 4 GHz, and 30 GHz to study low, mid, and high bands (as recommended by the 3GPP in TR 38.913 [31]) and base station antenna arrays of size 2x2, 4x4, and 8x8, respectively. Further parameters, in compliance with the aforementioned reports, are as follows: (a) base stations: i. 25 m high and ii. 49 dBm transmit power; (b) mobile user: i. 1.5 m high, ii. 0 dBi omnidirectional antenna, and iii. 9 dB noise figure; (c) -90 dBm noise floor (we assume noise to be independent of bandwidth).

2.3 What are the minimum spectrum requirements for the mobile network to meet URLLC goals?

Let us consider the case of $\delta = 32$ bytes and $\tau = 1$ ms [31]. We assume processing and medium access delays to be negligible. In practice, processing and medium access delays consume a fraction of the delay budget τ; however, these delays can be considered deterministic in ultra-reliable communication, as discussed in Section 2.2.1. We focus on the transmission delay, for it involves the primary sources of uncertainty in wireless networks, meaning that the transmission budget takes the entire latency budget, i.e., $\tau_{\text{tx}} = \tau$ in Equation (2.2). This allows us to quantify the *minimum* amount of spectrum needed to support ultra-reliable communication services.

We consider two downlink mobile communication scenarios. The first scenario corresponds to mobile networks subject to stochastic path loss and fading, and negligible interference, which we refer to as a noise-limited network, reflecting networks with a low frequency reuse or tight coordination between transmitters. The second scenario, an interference-limited network, includes interference, representing mobile networks subject to full frequency reuse, with little or no interference coordination between transmitters. Practical networks may operate somewhere in between noise- and interference-limited models, for they may adopt some forms of interference mitigation. Nevertheless, we consider noise- and interference-limited networks as the *lower*- and *upper*-bound estimates of the *fundamental minimum* amount of spectrum necessary to enable ultra-reliable communication services in mobile networks.

Our results stem from system-level Monte Carlo experiments. In each experiment, we consider the performance of a typical mobile user placed at the origin and base stations deployed according to a PPP of density λ BSs per km^2. The typical mobile associates with the base station of the highest average received power. We assume perfect beam alignment between mobiles and serving base stations and a fixed network load of one active connection per base station. This way, interference at the typical mobile is caused by the beam alignment between non-serving base stations and other randomly placed mobiles in the network. We generate 3×10^7 experiments for each network density to capture even low-probability channel conditions in the order of 0.001% when $\alpha = 99.999\%$.

2.3.1 Noise-limited networks

Figure 2.2 shows the demand for bandwidth to reach certain levels of reliability in different network densities for the three frequency bands of interest, 700 MHz, 4 GHz, and 30 GHz.

Let us consider a mobile network of density 30 BS/km^2, a density analogous to dense urban networks in [31], at the 30 GHz band as a reference. This network can be tailored to support reliable (i.e., $\alpha = 90\%$) to ultra-reliable communication (i.e., $\alpha = 99.999\%$) by the provision of an order of magnitude additional bandwidth (see vertical red arrows in Figure 2.2). Alternatively, increased density can ease the demand for bandwidth, such as illustrated by the horizontal red arrow in Figure 2.2, where ultra-reliable communication at $\alpha = 99.999\%$ demands as much bandwidth as less reliable regimes at $\alpha = 99.9\%$ at the cost of a roughly 2.3× denser deployment.

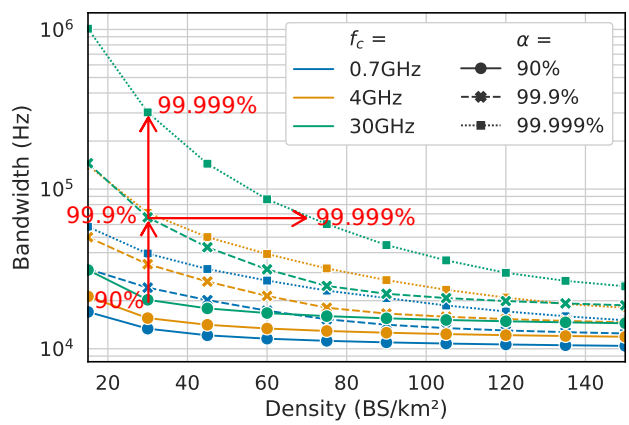


Figure 2.2: The bandwidth required to meet the ultra-reliable communication requirement of $\delta = 32$ bytes and $\tau = 1$ ms in noise-limited networks operating in different frequency bands.

The relationship between bandwidth and network density changes for different reliability levels α. An increase in network density can be traded for up to an order of magnitude of bandwidth in the ultra-reliability regime $\alpha = 99.999\%$ (dotted lines), whereas it has marginal impact on less reliable scenarios, such as $\alpha = 90\%$ (continuous lines). This is a consequence of how network density impacts coverage. As illustrated in Figure 2.1, the farther the mobile is from the base station, the harder it is to meet the reliability requirements because of weak coverage, increasing the demand for additional bandwidth. Increased network density improves coverage by reducing the distance between transmitter and receiver, as well as

the probability of non-line-of-sight communication. As ultra-reliability regimes are more sensitive to long-distance communication than less reliable regimes, they benefit the most from increased network density.

2.3.2 Interference-limited networks

Interference-limited networks demand additional spectrum to compensate for interference. Figure 2.3 is Figure 2.2's counterpart and shows that achieving $\alpha \geq 99.9\%$ can demand orders of magnitude more bandwidth than in noise-limited networks, whereas the requirements are somewhat similar for $\alpha = 90\%$ in both networks. Our system model assumes highly directional antennas with perfect beam alignment, which naturally decreases the overall interference, being it marginal when $\alpha \leq 90\%$. However, as the reliability requirement increases, communication is more sensitive to even low-probability interference conditions, such as in edge-case scenarios when nearby non-serving base stations (i.e., interferers) steer their beams towards the mobile, demanding wider bandwidth to compensate for strong interference.

The amount of bandwidth to support ultra-reliable communication is beyond what is typically available in today's networks. In dense urban networks, where network density is approximately 30 BS/km^2 [31], the required bandwidth to meet ultra-reliable communication at $\alpha = 99.999\%$ is around a gigahertz, which is impossible in the 700 MHz band and impractical in the 4 GHz band. The substantial amount of bandwidth favors the use of higher frequency bands, where wider bandwidth is typically available. Furthermore, higher frequency bands are often coupled with denser antenna arrays (as in our network model), enabling higher directionality transmission that reduces interference and increases resilience to path loss. This translates into less spectrum required to support ultra-reliable communi-

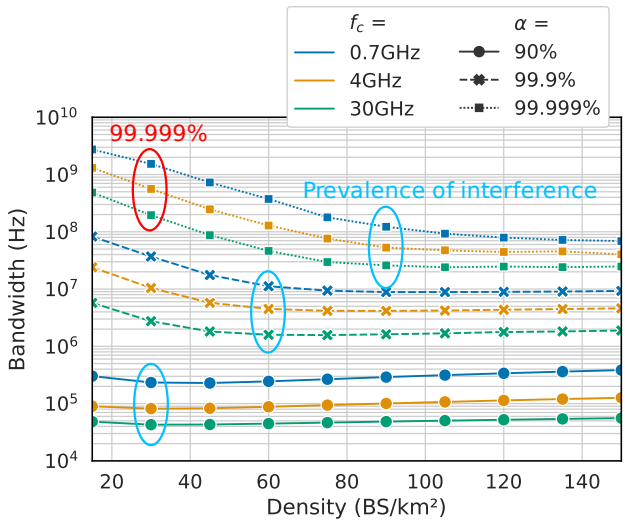


Figure 2.3: The bandwidth required to meet the ultra-reliable communication requirement of $\delta = 32$ bytes and $\tau = 1$ ms in interference-limited networks operating in different frequency bands. The red oval indicates the bandwidth required for ultra-reliable communication when the network density is 30 BS/km^2. Blue ovals indicate the network density where interference prevails for different reliability levels.

cation services than in the case of lower frequency bands.

Increased network density can ease the bandwidth requirements to some extent. Network density increases coverage until interference prevails, that is, dominant interfering links shift from non-line-of-sight to line-of-sight, and network densification starts to have a negligible (or even harmful) effect on coverage [33]. Interestingly, interference prevails at different network densities for different reliability levels. For instance, Figure 2.3 suggests that increased network density leads to marginal returns in less reliable regimes such as $\alpha \leq 90\%$, where interference prevails at approximately 30 BS/km^2. On the other hand, densification substantially eases the demand for bandwidth in higher reliability regimes such as $\alpha \geq 99.9\%$ up to the prevalence of interference at approximately 60 BS/km^2 and 90 BS/km^2, respectively.

Another way to look at the impact of network density is to consider the delay at a fixed bandwidth. Figure 2.4 illustrates the mapping from α and T_{tx} to network density when

bandwidth is fixed at 100 MHz in the 30 GHz band, the same amount of bandwidth recently auctioned in the US in auction 101 of the Federal Communications Commission (FCC) for one block of spectrum in upper 30 GHz bands. Lower reliability regimes (e.g., $\alpha \leq 99\%$) are only marginally impacted by network density; in fact, densification tends to increase the transmission delay in that region, as shown in the upper left side Figure 2.4. On the other hand, as we increase the reliability requirements (e.g., $\alpha \geq 99.999\%$), there is a substantial demand for increased network density. For instance, a general ultra-reliable communication requirement of $\alpha = 99.999\%$ and $\tau = 1$ ms demands a network density of 60 BS/km^2 or denser in Figure 2.4.

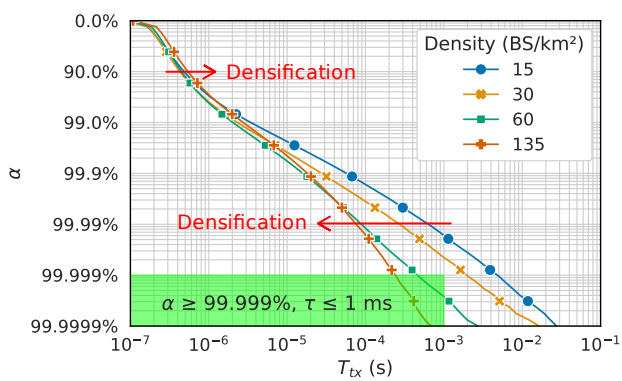


Figure 2.4: The impact of network density on the transmission delay when bandwidth is fixed at 100 MHz in the 30 GHz band. A general ultra-reliable communication requirement of $\alpha = 99.999\%$ and $\tau = 1$ ms (green zone) demands densities ≥ 60 BS/km^2. Similar patterns were found for the 700 MHz and 4 GHz bands.

The distinct trends in network density for different reliability regimes are due to how network density impacts overall and edge coverage. Unlike in noise-limited networks, where network density directly maps onto signal strength, the relationship between network density and coverage is subtle in interference-limited networks, for increased network density increases both signal strength and interference. As shown in Figure 2.5, increased network density incurs a small penalty for the overall coverage (i.e., upper right-hand side quadrant of Figure 2.5); on the other hand, increased network density significantly increases edge cov-

erage (i.e., lower left-hand side quadrant of Figure 2.5). (Results consistent with this finding are reported in [34]). The sensitivity to highly improbable events, such as edge conditions, increases with the reliability degree α (e.g., events with probability as low as 0.001% when $\alpha = 99.999\%$). As such, ultra-reliable communication is more sensitive to edge coverage than less reliable regimes, benefiting the most from increased network density.

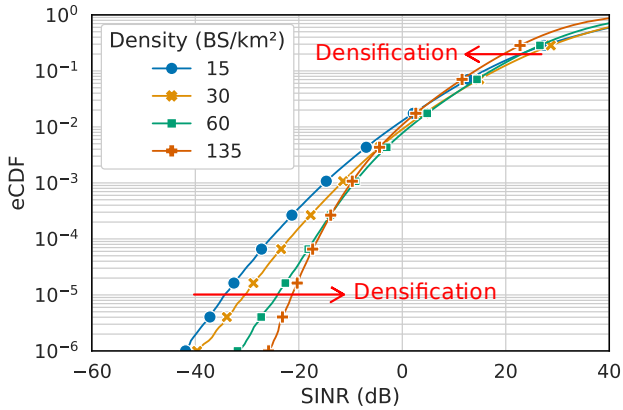


Figure 2.5: The empirical CDF of the SINR for different densities in the 30 GHz band. Similar patterns were found for the 700 MHz and 4 GHz bands.

2.3.3 Easing the bandwidth requirements

As discussed in the previous section, deploying mobile networks capable of supporting ultra-reliable communication at $\alpha \geq 99.999\%$ can demand massive amounts of resources (e.g., ≥ 1 GHz bandwidth or ≈ 100 BS/km^2). In this section, we consider multi-connectivity and multi-operator connectivity, which combines multi-connectivity with network sharing, as strategies to reduce this demand.

Multi-connectivity is an enabler of ultra-reliable communication that leverages diversity and redundancy in the form of multiple connections to mitigate the volatility of a wireless link [7, 8, 9]. In multi-connectivity, a mobile multi-connects to distinct base stations. The received signals can be combined in several ways using diversity-combining techniques. In

our analysis, we assume signals are weighted and summed according to the respective signal strengths, as in maximum ratio combining.

Multi-operator connectivity combines multi-connectivity and network sharing. In multi-operator connectivity, the mobile can multi-connect to base stations of different mobile operators in a multi-operator network. The multi-operator network sharing arrangement is akin to mobile virtual network operators like Google Fi, in the US, which operates atop several mobile operators. However, unlike Google Fi-like multi-operator networks, where users are often subject to the same radio resource management policy as subscribers of the underlying operators, we assume that bandwidth can be reserved and isolated for ultra-reliable communication users at each operator, for instance through network slicing. Furthermore, we assume users can simultaneously connect to base stations of more than one operator, as opposed to a single connection to a single base station at a time.

The resources in multi-operator networks differ from traditional single-operator networks in two ways. First, the base stations exhibit spatial correlation because of clustering between the involved operators [35]. Intuitively, clustering is caused by similar network planning strategies by the operators, such as covering areas of mutual interest, such as dense central business districts, thus deploying base stations adjacent to each other, or sharing towers, hence co-locating infrastructure. Second, multi-operator connectivity accounts for greater spectrum diversity because each mobile operator operates in its own licensed spectrum.

As in Section 2.3.2, we consider interference-limited networks. Figure 2.6 shows the bandwidth required to meet the ultra-reliable communication requirement of $\delta = 32$ bytes, $\tau = 1$ ms, and $\alpha = 99.999\%$. We focus our analysis on the 30 GHz frequency band; nevertheless, we can report that similar patterns were found in the 700 MHz and 4 GHz bands. Here, single-connectivity consists of a single connection to a single operator (as in previous sections;

hence our baseline); multi-connectivity, two connections, each to a base station of the same operator; and multi-operator connectivity, two connections, one connection to a base station of operator A and one connection to a base station of operator B. We model clustering in multi-operator connectivity as a Gauss-Poisson point process where clusters are deployed according to a PPP of density $\lambda/2$ [36]. Each cluster consists of two points: the first is at the center of the cluster and corresponds to a base station of operator A, and the second is randomly placed within a radius of 50 m and corresponds to a base station of operator B. Each mobile operator operates in its own licensed band with full frequency reuse.

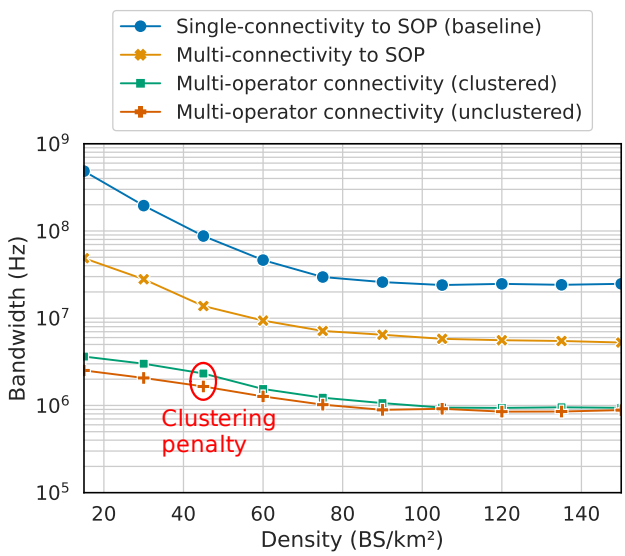


Figure 2.6: The bandwidth required to meet the ultra-reliable communication requirement of $\delta = 32$ bytes, $\tau = 1$ ms, and $\alpha = 99.999\%$ at the 30 GHz band in interference-limited networks. SOP stands for single-operator networks. The multi-operator network in multi-operator connectivity consists of two operators.

Multi-connectivity reduces the required bandwidth by approximately one order of magnitude. The reduction in the amount of bandwidth stems from the spatial diversity introduced by simultaneous connections to two base stations, which results in connections with distinct line-of-sight, path loss, and fading patterns. On the downside, both connections are subject to the same interference component because of full frequency reuse by the operator.

Multi-operator connectivity, in turn, benefits from frequency diversity, as each operator operates in its own licensed spectrum. As a result, multi-operator connectivity further reduces the required bandwidth, approximately two orders of magnitude compared to the baseline, from hundreds of megahertz to a few megahertz when the network density is 30 BS/km^2. Clustering incurs a small penalty in multi-operator connectivity compared to what it would be if the operators were unclustered (i.e., independent deployments), marginally increasing the demand for bandwidth. Furthermore, multi-operator connectivity is somewhat insensitive to the network density compared to single-connectivity and multi-connectivity and does not require massive network densification to reduce the demand for bandwidth, hence potentially reducing the cost of deploying networks capable of supporting ultra-reliable communication services.

2.4 Chapter summary

This chapter addressed spectrum dimensioning for ultra-reliable communication, a key step to enabling ultra-reliable communication services in future mobile networks. To that end, we mapped the ultra-reliable communication requirements onto bandwidth and assessed the amount of bandwidth it takes to support ultra-reliable communication services in large-scale 3GPP mobile networks. The amount of bandwidth depends on the reliability requirement of interest and can be in the order of gigahertz in ultra-reliability regimes such as $\alpha = 99.999\%$. Higher frequency bands, often coupled with wider bandwidth and denser antenna arrays than lower bands, are required in order to meet ultra-reliability goals. The demand for bandwidth can be traded for network density, but it requires massive densification (e.g., ≈ 100 BS/km^2) to significantly lower the required bandwidth, which is costly or even impractical, depending on the space available for the deployment of base stations. Alternatively, multi-operator

connectivity, a form of multi-connectivity to multiple mobile operators, was shown to reduce the demand for bandwidth by a few orders of magnitude without the outright deployment of a massive number of base stations.

Chapter 3

Multi-operator connectivity sharing for reliable networks: A data-driven risk analysis

3.1 Introduction

Reliability usually requires additional/redundant network resources, such as spectrum, antennas, or denser network deployments to compensate for the stochastic nature of wireless communications, also bringing additional CAPEX and OPEX for mobile operators. Network sharing is an alternative for mobile operators to mitigate upfront investments while leveraging additional network resources. Furthermore, a parallel trend for enhanced reliability in wireless communications is multi-connectivity, where mobiles simultaneously connect to several base stations, enhancing network reliability because of parallel, redundant connections [8, 9].

In this chapter, we advance our study of multi-operator connectivity sharing as an alternative for mobile operators to enhance network reliability while mitigating upfront investments in network infrastructure. Multi-operator connectivity sharing, where a mobile multi-connects to base stations of operators in the sharing agreement, is motivated by the increasing number

of mobile devices equipped with multi- and embedded-SIM cards, as well as initiatives to standardize multi-connectivity since 3GPP release 12. Furthermore, combinations of deployments by different operators exhibit characteristics that may provide additional support for reliability, such as increased spectral and spatial diversity, for underlying operators generally operate in their own licensed bands and infrastructure. In this chapter, we conduct an empirical evaluation of multi-operator connectivity sharing using a real-world dataset collected in the city of Dublin, Ireland. Our analysis indicates that multi-operator connectivity sharing benefits from complementarity between mobile operators, such as complementary coverage during periods of impaired connectivity, which naturally avoids service outages. However, additional connections incurred by multi-connectivity come at the cost of increased network load, which can increase the demand for resources (e.g., bandwidth in FDMA), potentially decreasing the capacity operators can offer to subscribers. Our study reports on the associated trade-offs and provides a set of lessons on how operators can leverage multi-operator connectivity sharing for reliable communication while minimizing the loss of aggregate capacity.

3.1.1 Related work

Multi-connectivity is an alternative approach to realize reliable communication – for comprehensive surveys, see [7, 25]. ReMP TCP [37] extends the multipath TCP (MPTCP) [38] to support reliable communication through the duplication of packets over multiple connections. The proposed approach suppresses the tail of the latency distribution to a greater extent than MPTCP and TCP in both system-level simulations and experimental evaluation of dual connectivity in an LTE mobile operator. In [9], the authors leverage diverse radio technologies (namely WiFi, 2G, 3G, and 4G) to enhance network reliability. They

explore several approaches to encode and split information over multiple connections. In practical scenarios, packet duplication significantly reduces latency while enhancing reliable communication, outperforming other more sophisticated packet-splitting strategies. In [8], the authors implement multi-base station connectivity for reliable communication in a small-scale testbed. Their approach requires coherent signal aggregation at the PHY layer, which makes it challenging to implement in large-scale real-world scenarios because of stringent synchronization requirements between base stations. In [39], the authors study multi-user networks in multi-connectivity scenarios. They propose a matching algorithm to assign secondary connections so as to optimize the data rate subject to minimum performance requirements. The authors of [40, 41] deploy multi-connectivity in two use cases, respectively: (a) an experimental testbed at the Hamburg seaport where multi-connectivity enhances network reliability by suppressing disruptions during handovers for mobile barges equipped with sensing devices; and (b) a small-scale industrial hall where multi-connectivity decreases the outage probability for mobile devices. [42] focuses on the use of multi-connectivity in wireless LAN, whereas [43] discusses architectural enhancements of LTE and 5G New Radio for multi-connected mobiles, both in light of reliable communication.

Multi-connectivity often requires additional resources to support multiple connections (be it antennas, spectrum, network density, etc.), which necessitates additional investments by mobile operators. Network sharing is an alternative way that mobile operators can leverage additional network resources while restricting capital and operational expenditures. The gains often come from the complementarity of multiple operators: there is no point in sharing resources if load and deployment patterns are the same [29, 44]. The analysis in [29] indicates that combining two operators can increase effective capacity in a cost-effective manner. The authors of [44] compare different forms of network sharing: (a) capacity sharing, modeled as a roaming process between two mobile operators; (b) spectrum sharing, where bandwidth

surplus is redirected to heavily loaded base stations; and (c) virtualized sharing, where spectrum is locally shared within mobile virtual networks. Their findings report on the effectiveness of capacity sharing, which performs better and is simpler to implement than the other forms of sharing that they studied. In [28], the potential of infrastructure and spectrum sharing is studied in the light of specific spatial characteristics of multi-operator networks, such as clustering, which is often found in real-world deployments [35]. On the operational side of network sharing, [26] shows that mobile operators can achieve significant energy savings by jointly switching off base stations in multi-operator networks while guaranteeing quality of service (QoS) to users. Allowing operators to jointly manage their networks can also increase effectiveness by curbing idle capacity while minimizing associated expenditures [27]. The collaboration versus competition dilemma is an important question in network sharing economics, and [45] studies whether two mobile operators should share network infrastructure. The authors demonstrate that network sharing has the potential to yield gains in some situations even if a mobile operator has the power to suppress the other or under competition regulation.

On the one hand, the aforementioned papers address multi-connectivity for increased network reliability and network sharing for cost-effective provisioning of network resources. On the other hand, none explores the potential of both approaches being jointly deployed, as in multi-operator connectivity sharing. In this chapter, we study multi-operator connectivity sharing as an alternative for operators to enhance network reliability without upfront investments.

3.2 Preliminaries

3.2.1 Network model

We consider a network with M operators and N mobile users, where each mobile is embedded with multiple radio frequency front-ends, each of which can support simultaneous connection to a different mobile network operator. All the mobiles are connected to their native operator, and some are also allowed to multi-connect to some other operators in compliance with what their native operator permits, as exemplified in Figure 3.1. Packets are duplicated over multiple connections (if available) and aggregated at the PDCP layer in light of dual-connectivity by the 3GPP since release 12. Reliability is ultimately considered end-to-end; however, we focus on the radio access network because the wireless environment is often the most dynamic, limiting, and expensive component of mobile networks. We consider the downlink communication and assess the system's performance based on the effective link capacity per mobile. We assume each operator occupies its own licensed spectrum and thus operators do not interfere with each other. We assume network interfaces are independent hardware-based implementations with negligible overhead impact on network performance. The effective link capacity of a mobile n connected to a set of base stations, $\mathcal{B}_n$, is given by:

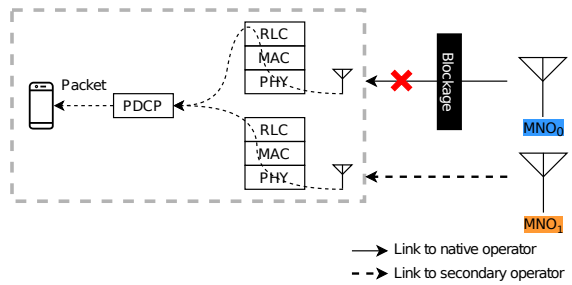


Figure 3.1: Example of multi-operator connectivity sharing with packet aggregation at the PDCP layer.

$$z_n = \max\{w_{n,b} \times \log_2(\gamma_{n,b} + 1), \forall b \in \mathcal{B}_n\} \text{ [bps]}, \tag{3.1}$$

where $w_{n,b}$ and $\gamma_{n,b}$ are the bandwidth and average SNR of the link between mobile n and base station b.

3.2.2 Network reliability

In this chapter, we define reliability as a measure of the α-superquantile of the downlink capacity and spectral efficiency. Superquantiles were proposed by [14, 15] as a measure of risk and are widely used in finance to assess low-probability losses. They are defined as the expectation over the lower tail of the density function of a random variable and depend on a parameter α that corresponds to the associated degree of risk. Let Z_m be a random variable corresponding to our figure of merit. Mathematically, superquantiles are defined as follows:

$$\bar{q}_\alpha(Z_m) = \frac{1}{\alpha}\int_0^\alpha q_\beta(Z_m)\mathrm{d}\beta, \tag{3.2}$$

where $q_\beta(Z_m)$ is the β-quantile of Z_m, i.e.,

$$q_\beta(Z_m) = \sup\{z \in \mathbb{R} : F_{Z_m}(z) \leq \beta\}, \tag{3.3}$$

where $F_{Z_m}(z)$ is the CDF of Z_m. Equivalently, if $F_{Z_m}(z)$ has no discontinuity, superquantiles can be written as $\bar{q}_\alpha(Z_m) = \mathbb{E}\left[Z_m | Z_m \leq q_\beta(Z_m)\right]$. One interesting mathematical property of superquantiles is that they approach the expected value when $\alpha \to 1$ and the infimum when $\alpha \to 0$, depicting their versatility.

3.2.3 Our data

We conducted walk tests and collected measurements from three mobile network operators in Dublin, Ireland (we will refer to them as MNO_0, MNO_1, and MNO_2). The resulting data

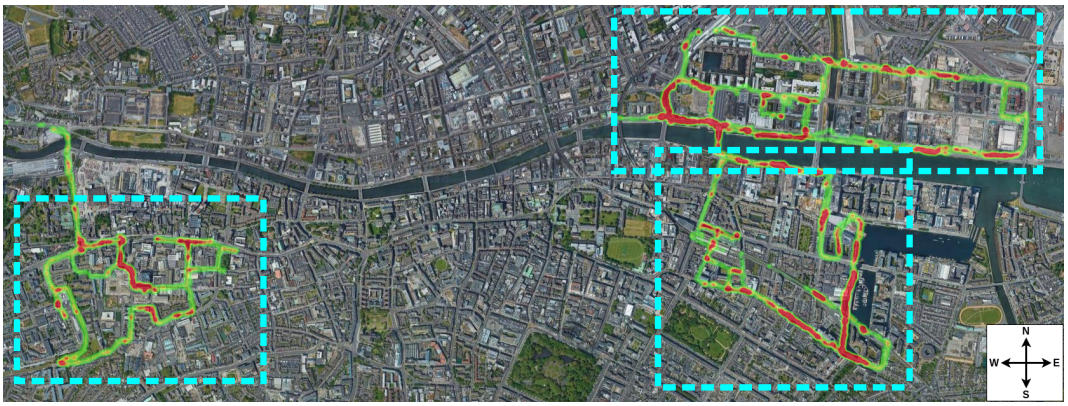


Figure 3.2: SNR traces of an operator for locations (from left-to-right, top-to-bottom): (a) North Dublin, (b) Guinness storehouse, and (c) South Dublin. Color legend: ● < -10dB, -10dB $\leq$ ● < 0dB, and 0dB $\leq$ ●.

includes timestamps, geographical coordinates, and performance metrics such as signal-to-noise ratio (SNR), reference signal received power (RSRP), and reference signal received quality (RSRQ) as shown in Table 3.1. These measurements were collected using G-MoN, a freeware passive observation application [46]. For a fair comparison, we used three LTE mobile phones (one per operator) of the same model and brand. The use of G-MoN involved activating the application to record data to a script file. During the walk tests, the phones were encased in a frame in a backpack to keep them at a similar orientation to each other so as to limit bias across the different operators. The phones recorded the data from the base station to which they were attached, once every second. In open areas of heavy traffic, the walkers would walk to a small number of separate points and pause for two to five minutes to gather a larger number of data samples. There was no downtime, i.e. the phones were always connected to a base station. Figure 3.2 depicts our traces over three different geographical areas of the city, namely North Dublin, the region around the Guinness Storehouse, and South Dublin.

Unique as they are, our traces have two main shortcomings in representing the network model of Section 3.2.1. First, they are limited to a single user per operator. We circumvent

Table 3.1: Illustrative example of our traces per operator at same times and locations.

Time	GPS coordinates	MNO_0	MNO_1	MNO_2
t_0	$coord_0$	snr_0,$rsrp_0$,$rsrq_0$	snr_0,$rsrp_0$,$rsrq_0$	snr_0,$rsrp_0$,$rsrq_0$
t_1	$coord_1$	snr_1,$rsrp_1$,$rsrq_1$	snr_1,$rsrp_1$,$rsrq_1$	snr_1,$rsrp_1$,$rsrq_1$
...	...	...	...	...
t_i	$coord_i$	snr_i,$rsrp_i$,$rsrq_i$	snr_i,$rsrp_i$,$rsrq_i$	snr_i,$rsrp_i$,$rsrq_i$

this limitation by considering that each entry in Table 3.1 could have been produced by a distinct user. The areas of heavy traffic have a higher density of data points. This mimics the geographical density of users and reflects locations where there is more demand for network resources. Second, there is no information on which base station each mobile was connected to at the time of data collection. In the absence of this information and operators' specific association policies, we assume that each user is connected to the geographically closest base station. To that end, we deploy a second public dataset on top of our traces [47], which contains the geographical coordinates of base stations of the three operators of interest. While in reality there may be cases when the mobile does not associate with the closest base station, sporadic mismatches do not significantly impact our analysis; our model and data capture the geographical demand of users for connectivity and the placement of network equipment to serve them. The resulting modified dataset is structured as shown in Table 3.2 and consists of 24763 possible users/locations and 40 base stations of MNO_0, 52 of MNO_1, and 56 of MNO_2.

Table 3.2: Illustrative example of the final dataset.

		MNO_0		MNO_1		MNO_2	
User id	Coord	BS id	SNR	BS id	SNR	BS id	SNR
u_0	$coord_0$	b_0	$\gamma_{0,0}$	b_1	$\gamma_{0,1}$	b_2	$\gamma_{0,2}$
...	...	...	...	...	...	...	...
u_j	$coord_j$	b_a	$\gamma_{j,a}$	b_b	$\gamma_{j,b}$	b_c	$\gamma_{j,c}$

3.3 Data analysis

In this section, we look at how multi-operator connectivity sharing impacts mobile communication. We assess the downlink capacity as a figure of merit for the three operators in our dataset. We contrast multi-operator connectivity sharing, or simply multi-operator connectivity (MC), with single-operator connectivity (SC) so that the gains and losses of multi-operator connectivity are made explicit. Subscripts indicate the operators in use. For example, we denote SC_0 to indicate that subscribers of MNO_0 operate in single-operator connectivity, MC_{01} to indicate that subscribers of MNO_0 and MNO_1 multi-connect to both operators' base stations (dual-connectivity), and MC_{012} to indicate that subscribers of the three operators multi-connect to base stations of each other (triple-connectivity). The total available bandwidth at each base station is set to 1 MHz for simplicity. We ran 100 Monte Carlo experiments for each network scenario (i.e. single-operator connectivity and variants of multi-operator connectivity), each experiment consisting of N_m active subscribers for each mobile operator m, where N_m is a Poisson random variable of mean λ_m. Let λ be the average density of simultaneously active mobiles per base station and k_m the number of base stations of MNO_m, such that $\lambda_m = \lambda k_m$. The N_m active users are randomly selected from our dataset and assigned to the operator's network for each Monte Carlo experiment. The placement of active mobiles is prone to overcrowding in a few base stations, for some locations have a higher density of data points as mentioned in Section 3.2.3. In real-world deployments, users would be eventually blocked once resources are finished until operators deploy more base stations in such areas to counteract overcrowding. In our experiments, we assume sufficient network infrastructure is already deployed to meet geographical demand, and therefore, subject each base station to a maximum of 100 active subscribers. This corresponds to the maximum number of resource blocks in an LTE system with 20 MHz bandwidth divided into resource blocks 180 kHz-wide and 10 kHz guard bands.

3.3.1 Coverage gains

Intuitively, the gains from multi-operator connectivity come from the complementary coverage and demand of different operators. That is, one operator may provide extra network resources when another fails. This is especially true for signal coverage because of the distinct placement of base stations, consequently leading to different signal propagation patterns. We reinforce this intuition by plotting the 10%-lowest SNR data points of a micro-region of our data in Figure 3.3. Each color refers to an operator: MNO_0 (red), MNO_1 (blue), and MNO_2 (yellow). In the area highlighted by the dashed circle, all three operators have weak signal coverage. Interestingly, however, the remainder of the data points indicate a high degree of complementary coverage between operators, such as pointed by the white arrow, where red predominates.

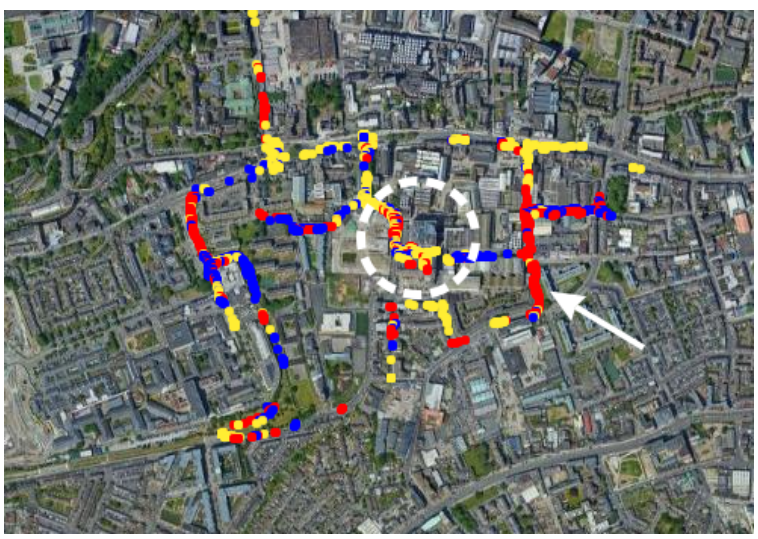
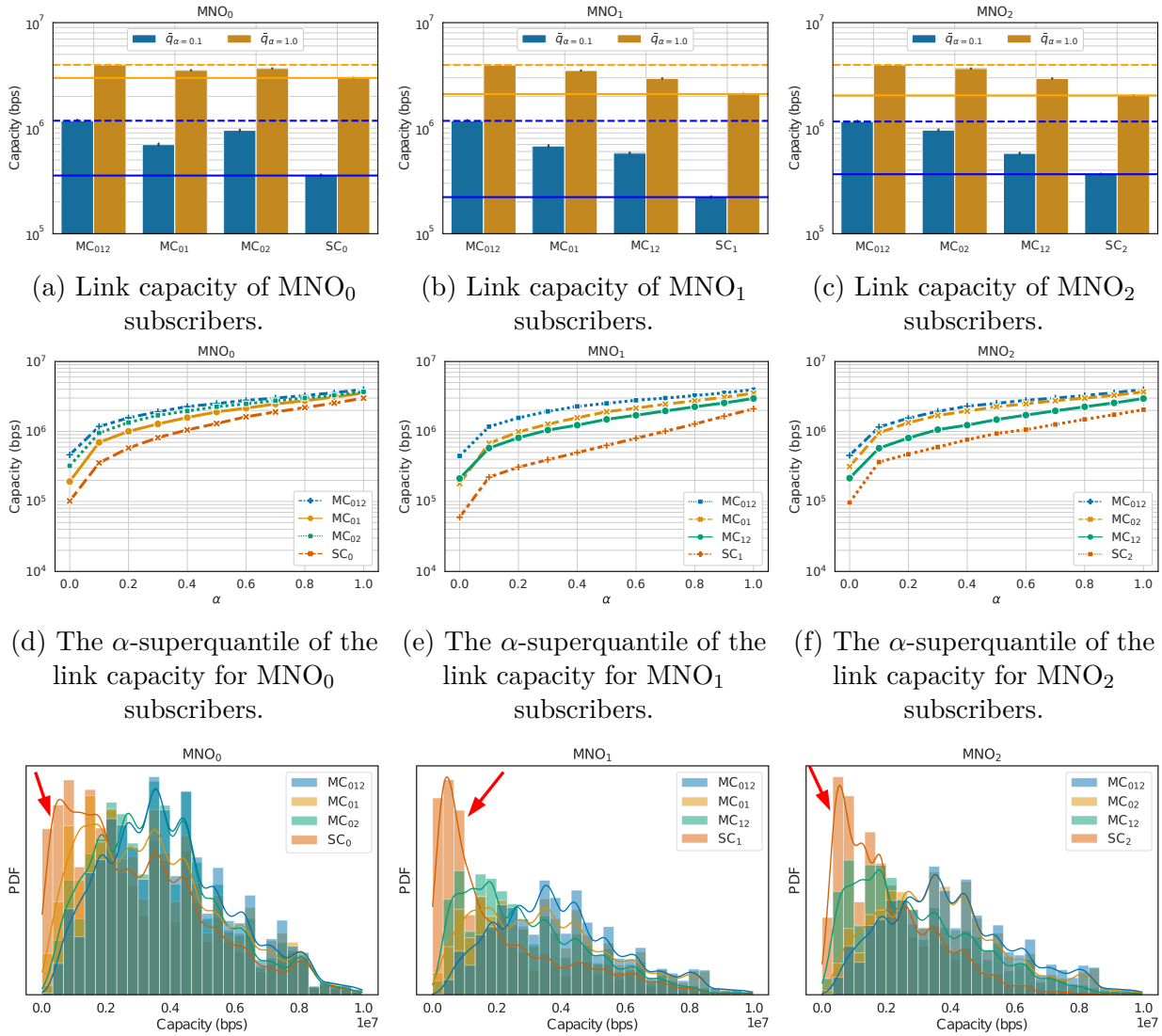


Figure 3.3: The lowest 10th percentile of SNR in the Guinness storehouse micro-region.

We start our analysis by evaluating the coverage gains of multi-operator connectivity. In MC_{ab}, all subscribers of MNO_a multi-connect to MNO_b's network and vice-versa. Let us consider the downlink capacity per mobile, a density λ of 10 active mobiles per base station, and the same unit bandwidth is allocated to all mobiles so that $w_{n,b} = 1$ MHz $\forall n \in N$ in

Equation (3.1).



(a) Link capacity of MNO$_0$ subscribers.

(b) Link capacity of MNO$_1$ subscribers.

(c) Link capacity of MNO$_2$ subscribers.

(d) The α-superquantile of the link capacity for MNO$_0$ subscribers.

(e) The α-superquantile of the link capacity for MNO$_1$ subscribers.

(f) The α-superquantile of the link capacity for MNO$_2$ subscribers.

(g) The PDF of the link capacity for MNO$_0$ subscribers.

(h) The PDF of the link capacity for MNO$_1$ subscribers.

(i) The PDF of the link capacity for MNO$_2$ subscribers.

Figure 3.4: The impact of multi-operator connectivity (MC) compared to single-operator connectivity (SC).

Figures 3.4a-3.4c show the 0.1-superquantile (blue bars) as well as the average capacity (orange bars) per operator. The interesting information is the relative gain from single-operator connectivity to multi-operator connectivity. The horizontal lines highlight this difference,

where continuous lines represent the performance of single-operator connectivity and dashed lines refer to the best performance of multi-operator connectivity. Operators benefit the most from multi-connectivity when all three operators participate in the shared network, i.e. MC_{012}. By contrasting the difference between continuous and dashed lines, we can clearly notice that the gains are much higher for the 0.1-superquantile than for the average user. Recall that the $\bar{q}_{\alpha=0.1}$ refers to the lower tail of the distribution of the capacity. The difference in gains decreases as we move from the lower-tail ($\alpha \to 0$) to the upper-tail ($\alpha \to 1$) as shown in Figures 3.4d-3.4f. This supports the application of multi-operator connectivity for reliable communication, as the gains mostly reside in situations where operators alone underperform. These results corroborate the observation from Figure 3.3 that operators tend to complement each other in scenarios of weak signal coverage.

The same results can be viewed from another angle as shown in Figures 3.4g-3.4i, the empirical probability density function (PDF)[1] of the capacity for the subscribers of each of the three operators, under single- and multi-operator connectivity. The most remarkable difference between these two connectivity schemes lies in the lower tail of the distributions (see red arrows). The smaller areas under the lower tails of the PDF for the case of multi-operator connectivity explain the greater gains in the 0.1-superquantile that we have previously discussed. This is intuitive, as operators often plan their network to provide overall good signal coverage in the same areas to supply their similar customers' demands, such as in high-density business locations (refer to [35] for clustering in multi-operator networks). The difference usually resides in parcels that are hard to plan for because, for example, of signal propagation effects or limitations given the location of the base stations.

To confirm this intuition, we explore the correlation of the coverage among operators. That

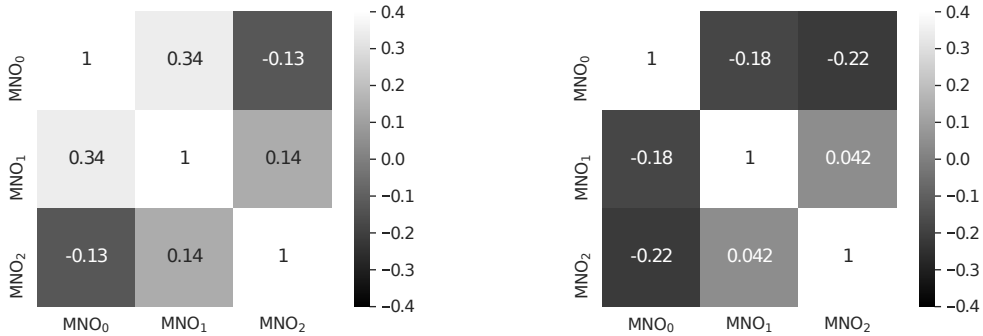
[1] The histograms correspond to the density of capacity measures from our experiments, whereas continuous lines are the kernel density estimate of the PDF.

is, if an operator has strong/weak signal coverage, how likely is it that neighboring operators also offer good/weak coverage? We pose this question with respect to all data points in our dataset and also regarding the intersection of the 10%-lowest SNR data points of each mobile operator. Figures 3.5a-3.5b show the Pearson correlation coefficients concerning the two cases. One can notice a significantly smaller correlation for the 10-percentile of the SNR data points, indicating that operators exhibit stronger complementarity in locations of weak signal strength than overall. A remark is that dual-connectivity combinations that lead to gains close to full multi-operator connectivity (i.e., MC_{012}) in Figures 3.4a-3.4c also present the smallest correlation coefficients, MC_{02} (i.e., dual-connectivity between MNO_0 and MNO_2), MC_{01}, and MC_{02}, respectively. This indicates, as we expect from intuition, that combining operators that strongly complement each other effectively yields coverage gains.

The results we have discussed rely on the assumption that each mobile in the network has the same fraction of the bandwidth $w_{n,b}$, making the capacity a function of the SNR only. This might relate to network deployments where spectrum is abundant. In reality, however, spectrum is often scarce and expensive, and the use of redundant connections in multi-operator connectivity tends to increase its demand.

3.3.2 Increased network load in multi-operator connectivity

In this section, we analyze a similar scenario to Section 3.3.1 but consider a frequency division multiple access (FDMA) network whereby the bandwidth available at each base station is allocated in a round-robin fashion and equally serves all the active mobiles, such that $w_{n,b} = 1/(\sum_{\forall n' \in \mathcal{N}_b} 1)$ MHz for all active mobiles connected to each base station b. Hence, the effective link capacity in Equation (3.1) depends on the number of active mobiles



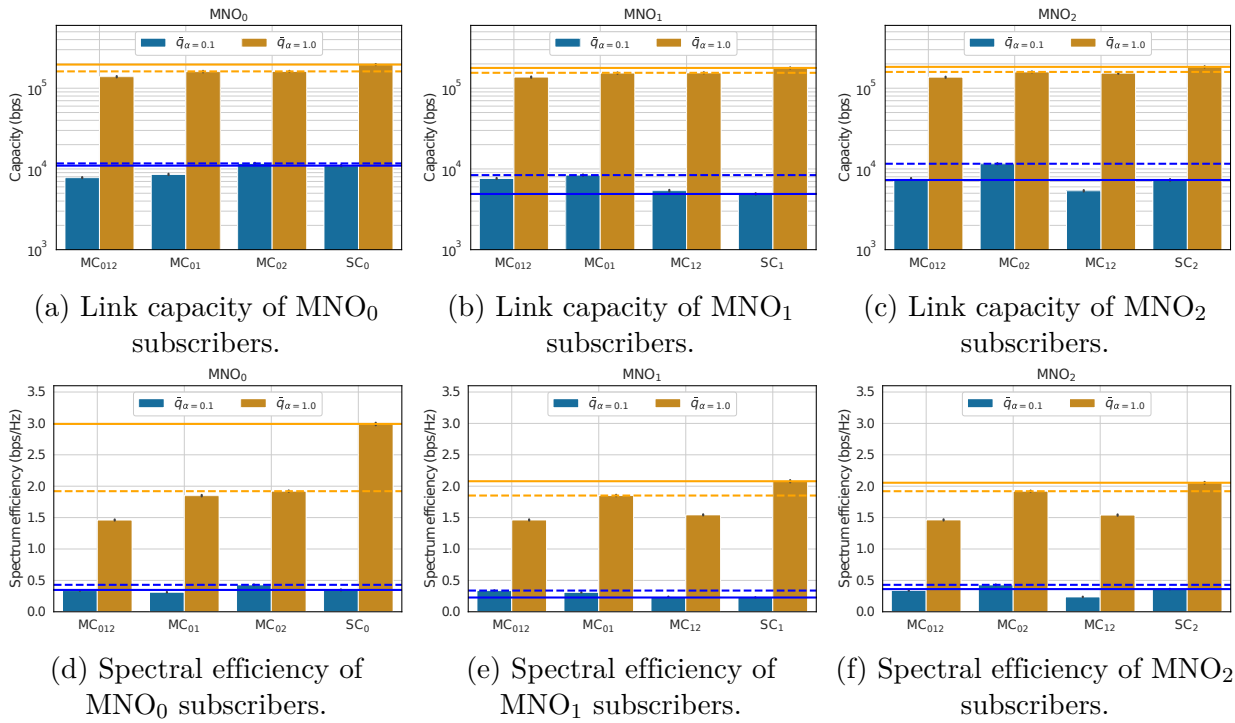
(a) All mobiles. (b) The 10-percentile.

Figure 3.5: Pearson correlation coefficients of SNR regarding all data points and the 10-percentile.

connected to each base station (i.e., network load) as well as on the SNR experienced by a mobile. Our goal is to capture the impact of increased network load incurred by multi-operator connectivity.

Figures 3.6a-3.6c are the counterparts of Figures 3.4a-3.4c and show the gains from single- to multi-operator connectivity in terms of the 0.1- and 1.0-superquantiles. Again, the horizontal lines contrast the best performance of multi-operator connectivity with single-operator connectivity for each operator. We see some reliability gains, especially for MNO_1, but the gains are modest in comparison with the previous case. This is even worse for the average capacity, which decays from single- to multi-operator connectivity, indicating a loss of performance.

In Figures 3.6d-3.6e, we further investigate such results by analyzing the cost of multi-operator connectivity. As redundant connections increase spectral use, we look into the spectral efficiency, the ratio of the effective link capacity to the total bandwidth allocated for each mobile n. Both single- and multi-operator connectivity lead to similar spectral efficiency in the worst case (i.e., $\bar{q}_{\alpha=0.1}$), whereas multi-operator connectivity is up to half as efficient as single-operator connectivity on average. The spectral efficiency observed by



(a) Link capacity of MNO_0 subscribers.

(b) Link capacity of MNO_1 subscribers.

(c) Link capacity of MNO_2 subscribers.

(d) Spectral efficiency of MNO_0 subscribers.

(e) Spectral efficiency of MNO_1 subscribers.

(f) Spectral efficiency of MNO_2 subscribers.

Figure 3.6: Link capacity and spectral efficiency in FDMA with limited bandwidth per base station. All mobiles multi-connect in multi-operator connectivity (MC). Mobiles only connect the their native operator in single-operator connectivity (SC).

mobile user n is calculated as:

$$\mathrm{s}_n = \frac{z_n}{\sum\limits_{\forall b \in \mathcal{B}_n} w_{n,b}} \text{ [bps/Hz]}. \tag{3.4}$$

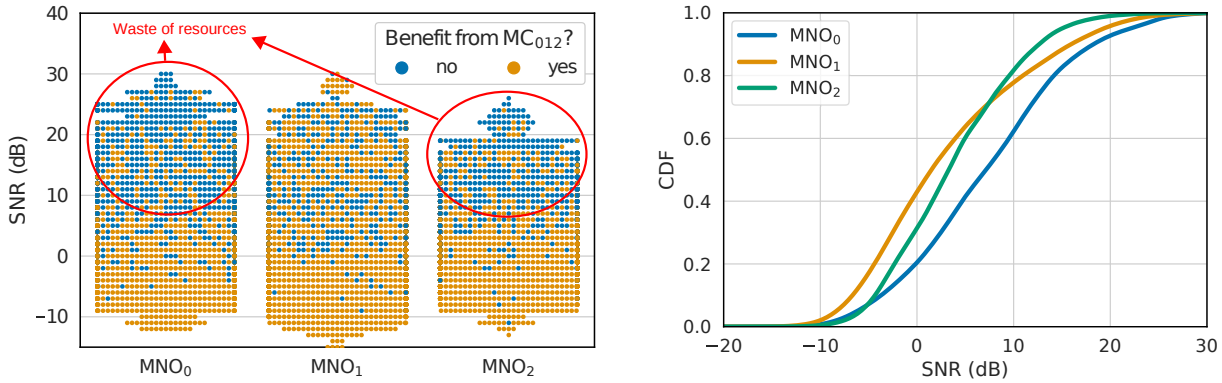
Recall that the link capacity is a linear function of the bandwidth and logarithmic in the SNR, so competition for spectrum can lead to lower capacity for individual mobiles, despite coverage gains. However, we have previously pointed out that the coverage gains are modest on average, implying that multi-operator connectivity may be ineffective in scenarios of high demand for network resources.

We take a step further in our analysis and quantify the benefit that multi-operator connec-

tivity brings to each user. To that end, we track the connection with the highest achievable capacity. If that connection is from the mobile's native operator, secondary links are not being used and are a waste of allocated bandwidth because the home operator already provides the best radio channel for communication. Otherwise, a mobile benefits from multi-operator connectivity, as secondary links outperform the primary connection.

In Figure 3.7a, each data point represents the SNR that a mobile experiences from its native operator. For ease of visualization, we randomly limit the number of data points presented. The orange dots represent users that have benefited from multi-operator connectivity (we refer to the benefit from full multi-operator connectivity, i.e., MC_{012}), whereas the blue dots those that have not. There are two remarkable facts that we can qualitatively observe here. First, many mobiles do not benefit from multi-operator connectivity – approx. 46% for MNO_0, 23% for MNO_1, and 33% for MNO_2. Second, the mobiles that benefit the most are the ones under weak signal coverage from their native operators – notice the higher density of orange dots for low values of SNR. The higher density of beneficiaries with high values of SNR for MNO_1 is because of its weaker signal coverage in comparison with others, lagging behind other operators in approximately 70% of our data as shown in Figure 3.7b. As a consequence, MNO_1 is more prone to link outages and a prominent candidate to benefit from multi-operator connectivity.

Motivated by our observation that mobiles experiencing low SNR are likely to benefit from multi-operator connectivity, we analyze a scenario where we only allow the subscribers of the (lower) β-percentile of SNR of each operator to multi-connect, where $\beta \in [0, 1]$. Those users are designated by their native operator and based on the instantaneous SNR of their native connection. The remaining subscribers are only granted a connection to their native operators. Let us first focus on the subscribers of MNO_0. Figures 3.8a and 3.8d show the 0.1- and 1.0-superquantiles of the link capacity, respectively. Multi-operator connectivity has a



(a) Strip plot of beneficiaries.

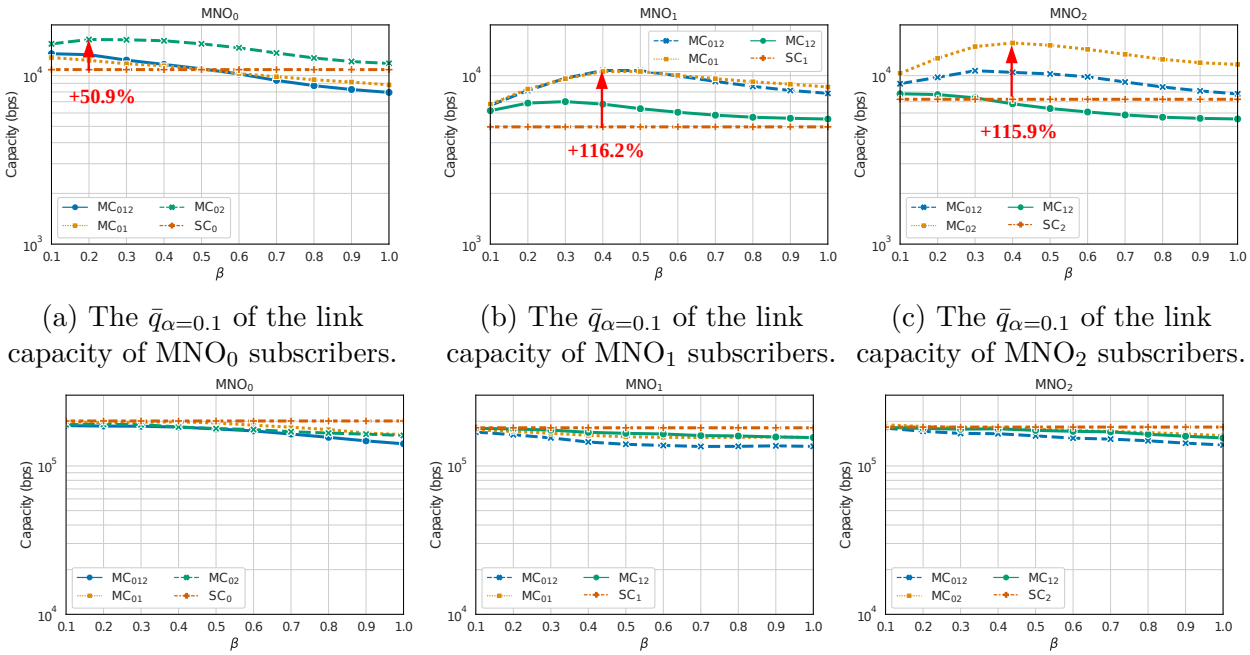
(b) The CDF of SNR.

Figure 3.7: Beneficiaries of multi-operator connectivity and empirical CDF of the SNR.

significant increase in reliability for small values of β (reaching up to 50.9% gains in MC_{02}) while having a low impact on the average performance. As more mobiles multi-connect ($\beta \to 1$), the gains proportionally diminish, impacting both 0.1- and 1.0-superquantiles of the link capacity.

These results stem from the impact of multi-operator connectivity on spectral efficiency. For small values of β, the spectral efficiency of multi-operator connectivity is even higher than single-operator connectivity as shown in Figures 3.9a and 3.9d. As we have previously pointed out, the users of low SNR are likely to benefit from multi-operator connectivity. However, the spectral efficiency decreases as mobiles that are already well covered by their native operator are allowed to also maintain redundant connections to other operators, leading to performance losses.

The same trends are also observed for MNO_1 and MNO_2, as shown in Figures 3.8b-3.8c and Figures 3.9b-3.9c. The only exception relates to subscribers of MNO_2 operating in multi-operator connectivity sharing with operator 1 (i.e., MC_{12}) in Figure 3.9c, which significantly underperforms single-operator connectivity (i.e., SC_2) for a wide range of values of β. Recall from Figure 3.5b that MNO_1 and MNO_2 are the only ones to exhibit positive correlation



(a) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_0 subscribers.

(b) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_1 subscribers.

(c) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_2 subscribers.

(d) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_0 subscribers.

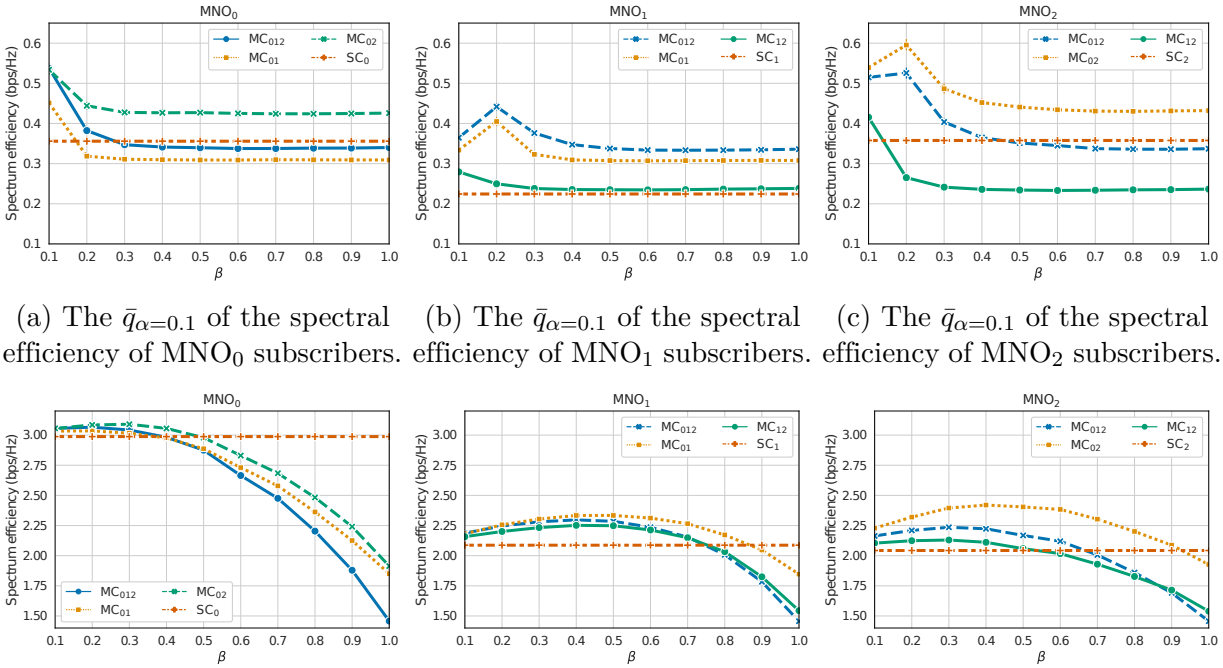
(e) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_1 subscribers.

(f) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_2 subscribers.

Figure 3.8: The 0.1- and 1.0-superquantiles of the link capacity. Only the subscribers of the lower β-percentile of SNR of each operator multi-connect in multi-operator connectivity (MC); the remaining users single-connect as in single-operator connectivity.

regarding their 10-percentile of SNR data points. The weak complementary coverage indicates that operators are unlikely to benefit from each other in such a case, explaining the lower spectral efficiency.

Interestingly, dual-connectivity is at least as good as full multi-operator connectivity (i.e., MC_{012}): MNO_0's subscribers benefit the most from MC_{02} in Figure 3.8a; MNO_1's subscribers similarly benefit the most from MC_{012} and MC_{01} in Figure 3.8b; and MNO_2's subscribers benefit the most from MC_{02} in Figure 3.8c. A similar trend is shown in Figures 3.8d-3.8f, where dual-connectivity incurs smaller penalties than MC_{012} on average. This is good news because, as we have previously discussed, multiple connections often come at the expense of lower spectrum efficiency. The best dual combinations (MC_{02}, MC_{01}, and MC_{02} in Fig-



(a) The $\bar{q}_{\alpha=0.1}$ of the spectral efficiency of MNO$_0$ subscribers. (b) The $\bar{q}_{\alpha=0.1}$ of the spectral efficiency of MNO$_1$ subscribers. (c) The $\bar{q}_{\alpha=0.1}$ of the spectral efficiency of MNO$_2$ subscribers.

(d) The $\bar{q}_{\alpha=1.0}$ of the spectral efficiency of MNO$_0$ subscribers. (e) The $\bar{q}_{\alpha=1.0}$ of the spectral efficiency of MNO$_1$ subscribers. (f) The $\bar{q}_{\alpha=1.0}$ of the spectral efficiency of MNO$_2$ subscribers.

Figure 3.9: The 0.1- and 1.0-superquantiles of the spectral efficiency. Only the subscribers of the lower β-percentile of SNR of each operator multi-connect in multi-operator connectivity (MC); the remaining users single-connect as in single-operator connectivity.

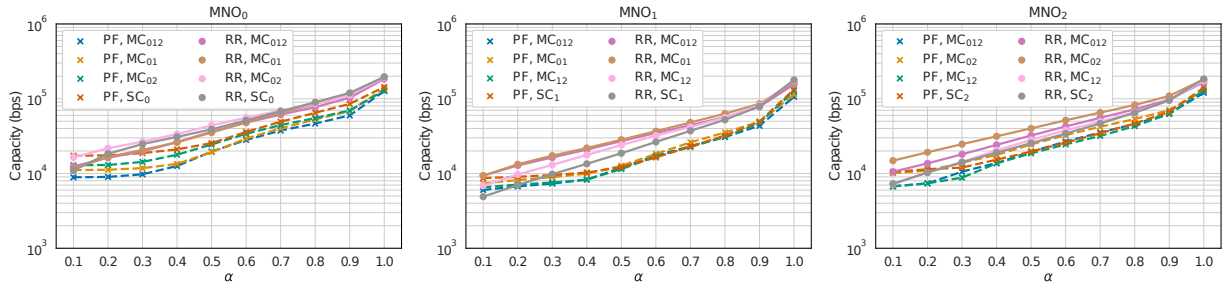
ures 3.8a-3.8c, respectively) coincide with small correlation coefficients in Figure 3.5b, also implying that multi-operator connectivity can achieve significant reliability gains with fewer redundant connections by selecting complementary mobile operators.

3.3.3 Scheduling and mobile density

So far we have analyzed settings where the bandwidth is allocated in a round-robin fashion. In this section, we also consider bandwidth assignment according to the proportional fair scheduling strategy, which assigns bandwidth to each mobile user n according to the metric $w_{n,b} = c_{n,b}/(\sum_{\forall n' \in b} c_{n',b})$ MHz where $c_{n,b} = 1/\log_2(\gamma_{n,b}+1)$ is the weight associated with the SNR. In this way, more bandwidth is assigned to active mobile users that experience low

SNR.

Figure 3.10 shows the α-superquantile of the link capacity of each mobile operator for round-robin and proportional as a function of α. We set $\beta = 0.3$, as it yielded significant gains in reliability with a small penalty in terms of spectral efficiency, as discussed in Section 3.3.2, and $\lambda = 10$ active mobiles per base station. Let us first focus on Figure 3.10a, which depicts the link capacity of subscribers of MNO_0. Therein, proportional fair with single-operator connectivity (i.e., PF, SC_0) outperforms round-robin with single-operator connectivity (i.e., RR, SC_0) for small values of α. This is expected since proportional fair favors mobile users of poor SNR, allocating more bandwidth to them. For $\alpha > 0.2$, in turn, round-robin tends to outperform proportional fair.



(a) Link capacity of MNO_0 subscribers.

(b) Link capacity of MNO_1 subscribers.

(c) Link capacity of MNO_2 subscribers.

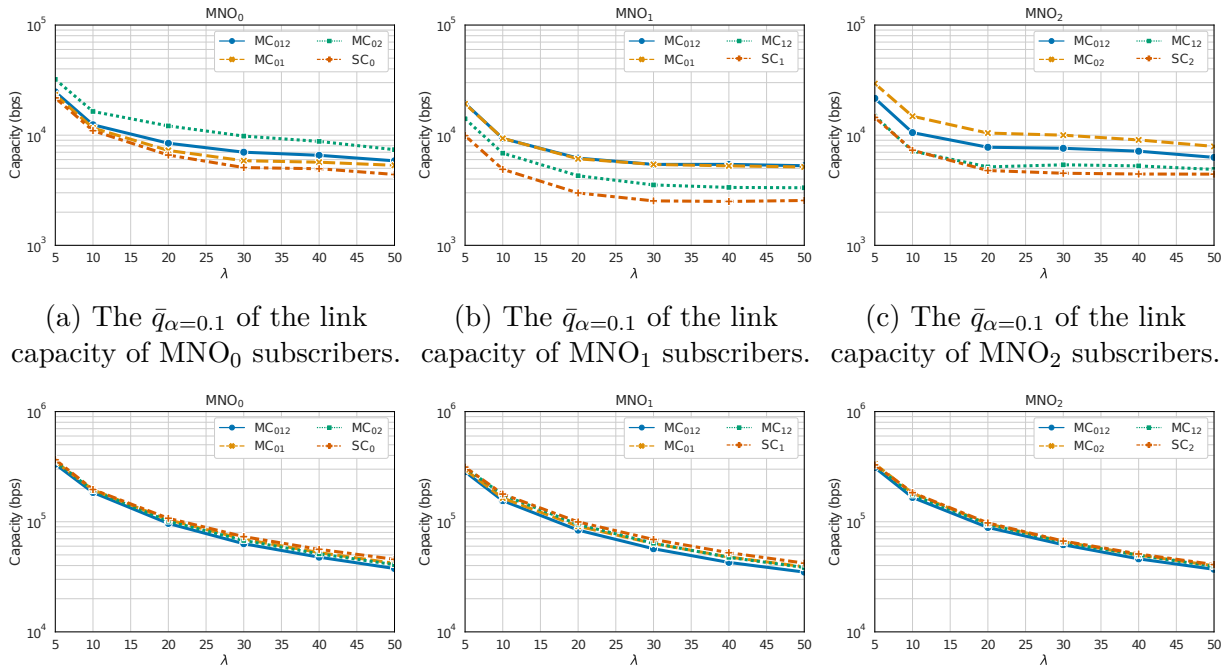
Figure 3.10: The α-superquantile of the link capacity for round-robing (RR) and proportional fair (PF) scheduling algorithms in single-operator connectivity (SC) and multi-operator connectivity (MC). $\beta = 0.3$, and $\lambda = 10$ active mobiles per base station.

Interestingly, proportional fair with multi-operator connectivity does not yield significantly higher gains than round-robin with multi-operator connectivity for small values of α. At first glance, we would expect a significant increase in reliability in proportional fair with multi-operator connectivity as both proportional fair and multi-operator connectivity tend to benefit under-performing users. However, it is noticeable that proportional fair and multi-operator connectivity are independently deployed in different network layers (MAC

and PDCP, respectively). We study this effect in more detail by considering conflicts between proportional fair and multi-operator connectivity. A conflict is when the interface of the highest allocated bandwidth (scheduled by proportional fair) does not coincide with the interface of the highest achievable capacity (selected by multi-operator connectivity). We denote conflict ratio as the percentage of mobiles that experience conflict between proportional fair and multi-operator connectivity. Here, the conflict ratios of MNO_0, MNO_1, and MNO_2 are approximately 50%, 70%, and 65%, respectively, for multi-connected mobiles in PF MC_{012} (i.e., proportional fair in full multi-operator connectivity). This result is intriguing as it indicates that conflicting network protocols can compromise (rather than jointly enhance) network reliability.

Lastly, we look at one more facet of multi-operator connectivity sharing, its dependence on the number of mobiles. As we have discussed in the preceding sections, increased network load can counterbalance the coverage gains of multi-operator connectivity. Recall that N_m, for all operators $m \in M$, is modeled as a truncated Poisson random variable dependent on the average density λ of simultaneously active mobiles per base station. Intuitively, this is an important aspect of multi-operator connectivity as the number of active mobiles in the network directly maps to the demand for spectrum. As we are about to see, however, it affects both single-operator connectivity and multi-operator connectivity to a similar extent.

Figures 3.11a and 3.11d show the results for the 0.1- and 1.0-superquantiles of the capacity as a function of the density of active mobiles of MNO_0. We set $\beta = 0.3$ and consider the round-robin scheduling strategy. As the density of active mobiles grows, so does the impact on the performance of single- and multi-operator connectivity. Nonetheless, the difference between them is mostly consistent over different mobile densities, and the gains and losses of multi-operator connectivity loosely depend on the number of active mobiles in the network.



(a) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_0 subscribers.

(b) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_1 subscribers.

(c) The $\bar{q}_{\alpha=0.1}$ of the link capacity of MNO_2 subscribers.

(d) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_0 subscribers.

(e) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_1 subscribers.

(f) The $\bar{q}_{\alpha=1.0}$ of the link capacity of MNO_2 subscribers.

Figure 3.11: The link capacity as a function of the average network density λ (active mobiles/base station). $\beta = 0.3$, and scheduling is round-robin.

Although we have constrained our attention to the subscribers of MNO_0, the same trends are observed for other mobile operators, as shown in Figures 3.10 and 3.11. The same applies to other values of $\beta \in [0, 1]$, and thus we have omitted these plots from this dissertation.

3.4 Chapter summary

This chapter advances our analysis of multi-operator connectivity sharing as a facilitator of ultra-reliable communication. To that end, we studied the trade-offs of multi-operator connectivity sharing using a real-world dataset of three mobile operators in the city of Dublin, Ireland. On the one hand, multi-operator connectivity sharing benefits from complementary coverage between operators, especially in locations of weak signal strength, where operators

tend to complement each other the most. On the other hand, the increased load incurred by multi-operator connectivity may throttle coverage gains and affect performance. To circumvent this drawback, mobile operators can increase reliability by focusing on the mobiles of weak SNR: a simple but effective strategy is only allowing weak SNR mobiles to multi-connect. As we have shown, this strategy significantly increases reliability at a small loss of average capacity. Increased network load can also occur because of an increase in the density of active users. This, however, affects single- and multi-operator connectivity similarly. Additionally, our results suggest that conflict between scheduling strategies and multi-operator connectivity can limit the gains of multi-operator connectivity sharing, requiring further investigation.

Chapter 4

Capturing rare network conditions to dimension resources for ultra-reliable communication

4.1 Introduction

A practical challenge of network dimensioning for ultra-reliable communication is the assessment of ultra-reliability regimes so that resources are dimensioned accordingly to compensate for even rare network conditions that may only occur with probability $\leq 0.001\%$ when reliability is $\geq 99.999\%$, as it is often the case for URLLC and HRLLC-like services. As an example, a rule of thumb for confident estimation is to have at least 100 likely instances of an event [48], implying that it requires the collection of $\geq 10^7$ network samples (or simulation cycles) to capture events that may only occur $\leq 0.001\%$ of the time, which can be impractical depending on the time it takes to collect each sample.

This chapter proposes an extreme value theory (EVT)-based approach to reduce the demand for data needed to dimension resources for ultra-reliable communication. Our contribution is threefold. First, inspired by our work in Chapter 2, we propose the use of the EVT framework to address the problem of the burden of massive data in spectrum dimensioning for

ultra-reliable communication, that is, how much data is needed to estimate the bandwidth required to meet the ultra-reliability requirements. Second, we extend the EVT framework by proposing the use of knee/elbow detection algorithms to automate one of the qualitative steps of the framework, as we will discuss in Section 4.2.2. Third, we observe that the proposed estimator can reduce the demand for data by several orders of magnitude compared to what would be needed otherwise. We show that the reduction in the demand for data is proportional to the reliability level, being of great relevance to ultra-reliable communication, where the desired reliability is $\geq 99.999\%$. Furthermore, the proposed estimator outperforms comparable state-of-the-art methods while also providing greater flexibility and generalizability.

4.1.1 Related work

Several approaches have been proposed to circumvent the burden of massive data in the context of ultra-reliable communication. One approach is to derive the distribution function of the random variable of interest (e.g., latency or signal strength), as in [49]. If the distribution is known, low-probability conditions can be assessed from the tail of the distribution. However, deriving the distribution function may only be practical for a handful of scenarios because of the complexity of network modeling (e.g., [49] omits interference and line-of-sight probability for tractability).

In the absence of the distribution function, an alternative approach is to rely on approximations of this function. For instance, [50, 51] used the kernel density estimation (KDE), a technique that approximates the probability density function of a random variable based on empirical data, to approximate the distribution of the SINR and then dimension the transmission rate for URLLC services. However, as we observe in Section 4.3, the KDE

does not approximate the tail of the distribution well and thus only marginally eases the demand for data in ultra-reliability regimes. In [52], the power-law approximation was applied directly to the tail distribution of the received power, and the obtained model was used to estimate low-probability conditions and dimension the transmission rate for URLLC services. This approach approximates well the tail distribution in a few stochastic network scenarios with different fading models. Yet, the power-law lacks generalizability and may produce inaccurate models, as we observe in Section 4.3.

Alternatively, EVT is shown to approximate well the tail of many distribution functions, being often used in risk analysis to model low-probability losses. The potential of EVT to enable ultra-reliable communication was suggested in [13]. Yet, to date, the application of EVT to ultra-reliable communication has been restricted to a handful of works, including [20], where EVT was adopted to approximate low-probability queue delays, and [21, 53], where EVT was used to approximate the tail distribution of the received power to select the transmission rate. To what extent EVT can be used to reduce the demand for data in ultra-reliability regimes remains an open question, which is the focus of this chapter.

4.2 Preliminaries

As in Section 2.2.1, this chapter adopts the ITU-R/3GPP definition of reliability: the success probability α of transmitting δ bits of data within a user plane latency deadline τ. We consider the user plane latency to be inversely proportional to the Shannon-Harley capacity:

$$T = \frac{\delta}{w \log_2(\Gamma + 1)}, \tag{4.1}$$

where w is the bandwidth available and Γ the SINR. This way, the reliability goal can be expressed as:

$$P\left(w \geq \frac{\delta}{\tau} S\right) \geq \alpha, \tag{4.2}$$

where $S = 1/\log_2(\Gamma + 1)$ is the inverse of the spectral efficiency. We assume the SINR to be constant in τ, a reasonable assumption for small values of τ. We also assume the SINR to be independent of the bandwidth, which implies that interference power dominates the noise power (i.e., the noise spectral density over a given bandwidth).

4.2.1 Spectrum dimensioning for ultra-reliable communication

The definition of reliability in Equation (4.2) implies that one way to achieve reliability is to ensure enough bandwidth to guarantee low-latency communication at the α-confidence level. This way, the minimum bandwidth to satisfy the service-level requirements δ, τ, and α can be expressed as the α-quantile of the random variable S:

$$w(\delta, \tau, \alpha) = \frac{\delta}{\tau} Q_S(\alpha), \tag{4.3}$$

where the α-quantile function is defined as:

$$Q_S(\alpha) = \min\{s : F_S(s) \geq \alpha\}, \tag{4.4}$$

with F_S denoting the cumulative distribution function of the inverse of the spectral efficiency. If F_S is known, we may calculate the bandwidth required to support ultra-reliable communication from Equation (4.4) and Equation (4.3). However, F_S will only be known for a handful of scenarios (as in [49]). Any more complex or practical network scenarios will likely require empirical evaluation of S based on a set of collected samples, where $\mathbb{S} = \{s_1, s_2, ..., s_n\}$ is the

set of n independently collected samples of the inverse of the spectral efficiency.

The direct approach to evaluate Equation (4.3) based on $\mathbb{S}$ is to approximate the value of $Q_S(\alpha)$ to the i-th element in $\mathbb{S}$, i.e., $\hat{Q}_S(\alpha) = s_i$, such that $i = \lceil \alpha \times n \rceil$, assuming $\mathbb{S}$ is sorted in ascending order (an approximation as in [54, Definition 1]). A shortcoming of this approach is its high demand for data when $\alpha \to 1$, as in ultra-reliable communication. If $n < 1/(1-\alpha)$, $Q_S(\alpha)$ will be approximated to the maximum value in $\mathbb{S}$, a crude approximation that generally leads to underestimation of the required bandwidth. A rule of thumb for confident approximation is to have $n \geq 100/(1-\alpha)$, so there are at least 100 samples greater than s_i (e.g., [48] recommends 120 samples for statistical inference), suggesting that $n \geq 10^7$ when $\alpha = 0.99999$ (i.e., 99.999% reliability), which can be impractical depending on the cost and time it takes to collect each network sample. We propose the use of EVT to estimate the spectral efficiency when the number of data samples is $n << 1/(1-\alpha)$.

4.2.2 Proposed EVT-based bandwidth estimator

The goal of EVT is to estimate rare events from a set of observations of a random variable of interest, being of great relevance when the number of observations is insufficient for a direct empirical approximation. The EVT methodology involves two general steps. The first step is to express the α-quantile of the random variable of interest S using the Pickands-Balkema-De Haan theorem. The second step consists of estimating the parameters of the obtained expression based on a set of observations $\mathbb{S} = \{s_1, s_2, ..., s_n\}$.

Let $X = S - u | S > u$ be the excess of S over a threshold u. The Pickands-Balkema-De Haan theorem postulates that the cumulative distribution function of exceedances tends to the generalized Pareto distribution (GPD) for a large class of distribution functions and sufficiently high threshold u, i.e., $X \sim$ GPD. This allows us to approximate high α-quantiles

of S with:

$$Q_S(\alpha) \approx \begin{cases} u - \beta \ln\left(\frac{1}{1-F_S(u)}\right) - \beta \ln(1-\alpha), & \xi = 0, \\ u + \frac{\beta}{\xi}\left(\left(\frac{1}{1-F_S(u)}(1-\alpha)\right)^{-\xi} - 1\right), & \xi \neq 0. \end{cases} \tag{4.5}$$

(The reader can refer to [55] for the derivation of Equation (4.5)).

The parameters ξ and β are the shape and scale parameters of the cumulative distribution function of X. The cumulative distribution function of X is known. Thus, we can obtain the parameters ξ and β by fitting the set of samples $\mathbb{X} = \{x_1, x_2, ..., x_{n_u}\}$ to the generalized Pareto distribution using the maximum likelihood estimator, where $\mathbb{X}$ is the set of exceedances over u, i.e., $\mathbb{X} = \{s - u, \forall s \in \mathbb{S} | s > u\}$; the maximum likelihood estimator algorithm to estimate the parameters of the generalized Pareto distribution is described in [56]. Furthermore, $F_S(u)$ can be numerically approximated with $(n - n_u)/n$, where n and n_u are the cardinalities of $\mathbb{S}$ and $\mathbb{X}$.

What remains is to select the threshold u. There are two common approaches to select u [57]. The first is to treat the threshold as fixed, such as setting u to the i-th element in $\mathbb{S}$, assuming $\mathbb{S}$ is sorted. However, this approach is insensitive to variations in subsequent sets of data, which can compromise the goodness of fit of exceedances to the generalized Pareto distribution and thus the quality of estimates. The second approach consists of a qualitative assessment of the goodness of fit of exceedances to the generalized Pareto distribution. For instance, an observed characteristic of the generalized Pareto distribution is the linearity of the mean excess function, i.e., $e(v) = \mathbb{E}(S - v | S > v)$, for a sufficiently high v. This way, the threshold u can be selected using plots of $e(v)$ and set as the point where $e(v)$ becomes linear for all $v > u$. However, using plots of the mean excess function requires an expert to assess the model fit, a rather manual and subjective step.

In our EVT-based estimator, we propose using knee/elbow detection algorithms, e.g., [58],

to automate the threshold selection leveraging the mean excess function. The knee/elbow detection algorithm tracks the curvature of the estimated mean excess function and selects the threshold u as the point where it becomes linear, i.e., the knee/elbow point. The reasoning for selecting the smallest threshold such that the mean excess function is linear is to maximize the cardinality of the set of exceedances, which, in turn, contributes to improving the estimates of the parameters ξ, β, and $F_S(u)$.

Algorithm 1 summarizes the proposed EVT-based bandwidth estimator. First, it obtains the threshold u using a knee/elbow detection algorithm. Then, it obtains the set of exceedances and estimates the shape and scale parameters of the generalized Pareto distribution using the maximum likelihood estimator, i.e., $\hat{\xi}$ and $\hat{\beta}$. Lines 6-10 correspond to the bandwidth estimation in Equation (4.3) with $Q_S(\alpha)$ as in Equation (4.5) and $F_S(u) \approx (n - n_u)/n$. As we show subsequently, we can use Algorithm 1 to estimate the required bandwidth for ultra-reliable communication even if $n << 1/(1-\alpha)$.

Algorithm 1 Proposed EVT-based bandwidth estimator

1: **function** EVT-ESTIMATOR$(\mathbb{S}, \delta, \tau, \alpha)$
2: $\quad u \leftarrow$ select the threshold using knee/elbow detection
3: $\quad \mathbb{X} \leftarrow \{s - u, \forall s \in \mathbb{S} | s > u\}$
4: $\quad \hat{\xi}, \hat{\beta} \leftarrow$ fit $\mathbb{X}$ to the GPD using MLE
5: $\quad n, n_u \leftarrow |\mathbb{S}|, |\mathbb{X}|$
6: $\quad$ **if** $\hat{\xi} \approx 0$ **then**
7: $\quad\quad w \leftarrow \frac{\delta}{\tau}\left(u - \hat{\beta}\ln\left(\frac{n}{n_u}\right) - \hat{\beta}\ln(1-\alpha)\right)$
8: $\quad$ **else**
9: $\quad\quad w \leftarrow \frac{\delta}{\tau}\left(u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{n_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right)\right)$
10: $\quad$ **end if**
11: $\quad$ **return** w
12: **end function**

4.2.3 State-of-the-art methods

Power law. The power-law approach is based on the findings in [52], which demonstrates that, for selected fading models and in the absence of interference, the tail of the received signal strength can be well approximated by $F_Y(y) \approx \zeta y^{\frac{1}{\kappa}}$, where ζ and κ are parameters of the distribution. This way, we define the α-quantile of the inverse of the spectral efficiency as:

$$Q_S(\alpha) = \frac{1}{\log_2\left(1 + Q_Y(1-\alpha)\right)}, \tag{4.6}$$

where, given $F_Y(y) \approx \zeta y^{\frac{1}{\kappa}}$,

$$Q_Y(\alpha) \approx \left(\frac{\alpha}{\zeta}\right)^{\kappa}. \tag{4.7}$$

Let $Z = \ln(Y)$ and $\mathcal{Z} = \{z_1, z_2, ..., z_n\}$, where $z_i = \ln(y_i)$ is an observation of the received signal strength. Following [52], the parameters ζ and κ can be obtained from set $\mathcal{Z}$ as in:

$$\hat{\kappa} = \frac{1}{l}\sum_{i=1}^{l}(z_l - z_i), \tag{4.8}$$

$$\hat{\zeta} = \frac{l}{n} e^{\frac{-z_l}{\kappa}}, \tag{4.9}$$

where $l = \lceil \beta n \rceil$ and $\beta = 0.01$. As in Equation (4.3), the required bandwidth can be obtained as $w(\delta, \tau, \alpha) \approx \frac{\delta}{\tau}\hat{Q}_S(\alpha)$.

Kernel density estimation (KDE). KDE algorithms are used to estimate the PDF of a random variable S based on a set of observations $\mathcal{S} = \{s_1, s_2, ..., s_n\}$,

$$\hat{f}_S(s) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{s - s_i}{h}\right), \tag{4.10}$$

where K is the kernel function and h a smoothing parameter. In this chapter, as in [50, 51], we

use (a) Gaussian kernels for K and (b) Scott's rule to obtain h [59] (a Python implementation of the KDE with Gaussian kernels can be found in [60]). This way, we can estimate the CDF of S,

$$\hat{F}_S(s) = \int_{-\infty}^{s} \hat{f}_S(x)\mathrm{d}x, \tag{4.11}$$

and the α-quantile of S,

$$\hat{Q}_S(\alpha) = \min\{s : \hat{F}_S(s) \geq \alpha\}, \tag{4.12}$$

allowing us to obtain the required bandwidth $w(\delta, \tau, \alpha) \approx \frac{\delta}{\tau}\hat{Q}_S(\alpha)$, as in Equation (4.3).

4.3 Numerical evaluation

We consider the 3GPP-URLLC service-level requirements of $\delta = 32$ bytes, $\tau = 1$ ms, and $\alpha = 0.99999$. Our goal is to quantify the amount of data needed to dimension the bandwidth required to support the ultra-reliable communication service.

We consider the spectrum dimensioning approaches discussed in Section 4.2. The estimator in Section 4.2.1 is our baseline to assess the efficiency of our solution proposed in Section 4.2.2. As the ground truth, we use the baseline estimator applied to a set of 2×10^8 network samples. This way, we can compare how much data each estimator takes to approximate the ground truth bandwidth.

To collect the data, we run simulations on two 3GPP network scenarios operating at the frequency bands of 4 GHz and 30 GHz. For each scenario, we collect n independent network samples to form $\mathbb{S} = \{s_1, s_2, ..., s_n\}$, where each sample is the inverse of the spectral efficiency experienced by a typical mobile located at the origin of a network deployed according to a Poisson point process with an expected density of 30 base stations per km^2 (a density typical

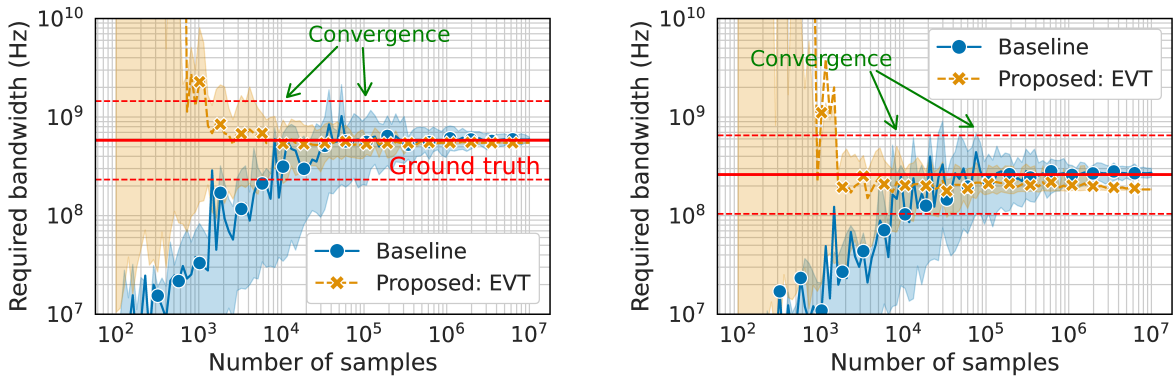
of dense urban network deployments [31]).

The network scenarios are in accordance with the urban macro (4 GHz band) and urban micro (30 GHz band) 3GPP models described in the 3GPP TR 38.901 [32], which defines antennas, path loss, fading, and line-of-sight probability. Base stations are equipped with {4x4, 8x8} antenna arrays, are {25, 10} m high, and have a {49, 35} dBm transmit power for frequency bands {4, 30} GHz. The networks are interference limited, meaning that, for each model, base stations operate at the same frequency band and are active all the time, with a fixed network load of one mobile per base station. Mobile handsets are 1.5 m from the ground level and equipped with a 0 dBi omnidirectional antenna. A mobile attaches to the base station that provides the highest average received power. The noise figure and noise floor are 9 dB and -90 dBm, respectively.

4.3.1 How much data is needed to dimension the spectrum for ultra-reliable communication?

Figure 4.1 shows the bandwidth estimates for each estimator as a function of the number of collected samples, n. We repeated the estimations thirty times for each value of n to produce confidence intervals, each of which with a different set $\mathbb{S} = \{s_1, s_2, ..., s_n\}$. The marked lines correspond to the expectation of the estimated bandwidth, whereas the shaded curves show the confidence interval between the 10-th and 90-th percentiles of the obtained estimations.

Of course, it is hard to infer the bandwidth at the 99.999% reliability level if data is scarce (e.g., $n = 10^2$), for data samples are unlikely to capture any asymptotic behavior of the network. The quality of estimations improves as more data is made available, with estimations tending to the ground truth (continuous red line) as the number of samples approaches



(a) 4 GHz band.

(b) 30 GHz band.

Figure 4.1: Bandwidth dimensioning at $\alpha = 0.99999$ as a function of the number of network samples.

10^7. Notice that the baseline estimator approximates the ground truth from the bottom up, with a tendency to underestimate the bandwidth. As discussed in Section 4.2.1, for a number of samples $n < 1/(1-\alpha)$, the baseline approximates the bandwidth based on the maximum value in $\mathbb{S}$, a crude approximation depending on how heavily tailed distribution of the underlying random variable S is.

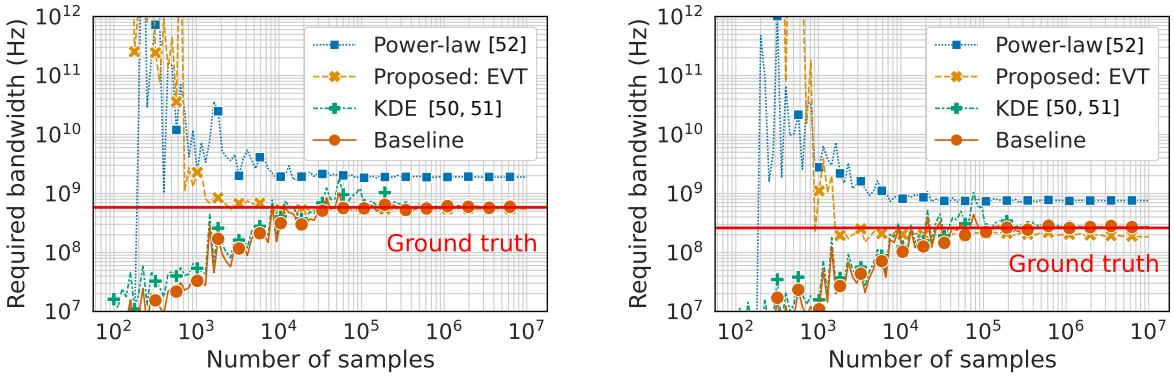
The exact point of convergence depends on the acceptable error. In Figure 4.1, we accept an error of 2.5× the ground truth (red dashed lines) so that convergence corresponds to confident estimations of the same order of magnitude of bandwidth as the ground truth. This way, the EVT estimator takes approximately 10^4 samples to converge, a cost reduction of one order of magnitude compared to the baseline and four orders of magnitude compared to the ground truth. Of course, the accepted error is dependent on the application. Nevertheless, Figure 4.1 indicates that the EVT estimator is consistently closer to the ground truth for $10^3 \leq n \leq 10^6$. The baseline estimator can indeed outperform the EVT estimator if the number of samples is massive (e.g., $n = 10^7$ in Figure 4.1b). Remember, however, that the main motivation for using the EVT estimator is to avoid the burden of collecting massive amounts of data.

4.3.2 Benchmark against state-of-the-art methods

In this section, we extend the evaluation of the proposed EVT estimator to two state-of-the-art approaches used in the context of ultra-reliable communication and presented in Section 4.2.3.

Figure 4.2 is Figure 4.1's counterpart and shows the expected bandwidth estimates for the two state-of-the-art approaches mentioned above as well as the baseline and the proposed EVT estimator. We omit the confidence intervals for ease of visualization. The proposed EVT estimator approaches the ground truth faster than its counterparts. The power-law estimator overestimates the amount of required bandwidth regardless of the number of samples – it does not even converge according to our convergence criterion in Section 4.3.1. We reason that the miscalculation stems from the fundamental nature of the power-law estimator. Unlike the EVT framework, which is shown to approximate the tail distribution of a large class of distribution functions, the power-law is only shown to approximate the tail distribution for selected fading models and in the absence of interference and obstacles [49, 52]. The KDE estimator performs similarly to the baseline, which is expected because the KDE is primarily a smoothing technique that approximates the probability density function of a random variable, as opposed to the tail of the distribution, as in EVT.

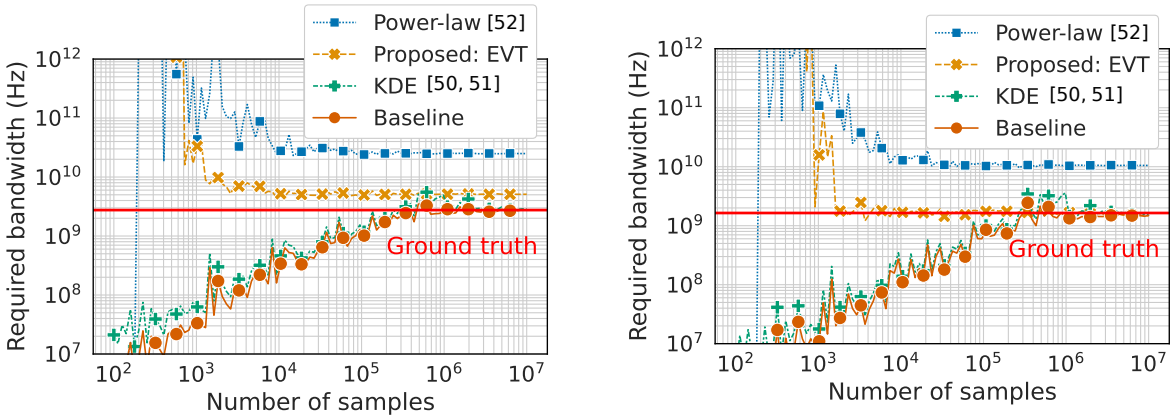
Furthermore, Figure 4.3 shows the performance of the estimators when $\alpha = 0.999999$, an additional "nine" compared to Figure 4.2. Again, the proposed EVT estimator outperforms its counterparts. Notice, however, that the gains, in the form of the number of data samples, are greater than what was shown before. The EVT estimator is fairly consistent in both reliability regimes. On the other hand, as the reliability requirement increases, so does the error by the power-law estimator and the number of samples needed for convergence by the KDE and baseline estimators, implying that the attractiveness of the EVT approach



(a) 4 GHz band. (b) 30 GHz band.

Figure 4.2: Benchmark when $\alpha = 0.99999$.

increases with the reliability requirement.



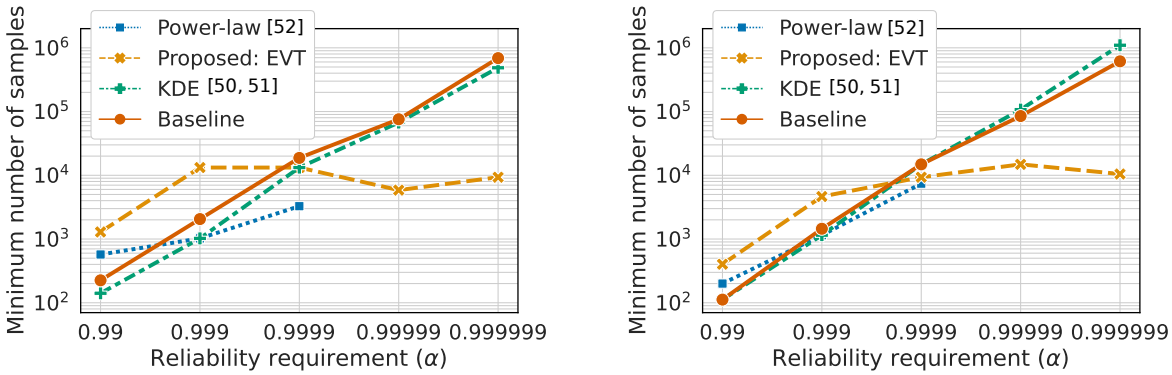
(a) 4 GHz band. (b) 30 GHz band.

Figure 4.3: Benchmark when $\alpha = 0.999999$.

4.3.3 The amount of data according to the reliability requirement

Intuitively, the higher the reliability requirement, the more challenging it is to dimension resources for ultra-reliable communication services. In this section, we study the number of samples needed to dimension the bandwidth as a function of the reliability requirement α. Using the same convergence criterion as in Section 4.3.1, i.e., confident estimations within

2.5× the ground truth, Figure 4.4 shows the minimum number of samples required for convergence as a function of the reliability requirement. The demand for data tends to increase as a function of the "nines". On the one hand, the EVT estimator demands more data than other estimators for lower reliability requirements; on the other hand, the trend changes after 0.9999, where the EVT estimator requires less data to converge. The increasing performance of the EVT estimator is expected because it approximates the asymptotic behavior of the spectral efficiency, i.e., $\alpha \to 1$, as pointed out in Section 4.2.2. Remarkably, the EVT estimator consistently converges with approximately 10^4 samples when $\alpha \geq 0.9999$, as opposed to the baseline and KDE estimators, which increase the demand for data by about one order of magnitude per "nine". In both data sets, the power-law estimator does not converge when $\alpha > 0.9999$ regardless of the number of data samples.



(a) 4 GHz band. (b) 30 GHz band.

Figure 4.4: Minimum number of samples required for convergence for different reliability requirements.

4.4 Chapter summary

This chapter addressed the challenge of massive data for spectrum dimensioning in ultra-reliable communication. We demonstrated that EVT can be an effective way to reduce

the demand for data while estimating the required bandwidth to meet stringent reliability requirements. The proposed EVT-based estimator outperformed conventional and state-of-the-art methods in 3GPP network scenarios. The gains are proportional to the reliability requirements. Despite modest (or even adverse) performance in lower reliability regimes, EVT was found to be of great relevance in ultra-reliability regimes, where reliability is often $\geq 99.999\%$, such as in many URLLC- and HRLLC-like services, reducing the demand for data by several orders of magnitude compared to what would be needed otherwise.

Chapter 5

Assessing the spectrum needs for network-wide ultra-reliable communication with meta distributions

5.1 Introduction

This chapter advances spectrum dimensioning for ultra-reliable communication services by proposing an analytical framework to capture the relationship between spectrum needs and service-level requirements of ultra-reliable communication services. The proposed framework, which is based on meta distributions and will be further discussed in Section 5.2, differentiates between two notions of reliability (see Figure 5.1). The link reliability (α) corresponds to the success rate of meeting the performance requirements of ultra-reliable communication services. The spatial reliability (η) corresponds to the service area, in percentages, in which ultra-reliable communication is supported. This chapter's main contribution is closed-form expressions for the required bandwidth to support the target reliability requirements α and η. The obtained expressions reveal a surprisingly simple relationship between spectrum needs

and ultra-reliability service-level requirements: each "nine" in reliability translates into an order of magnitude increase in the required bandwidth. Inspired by our previous works in Chapters 2 and 3, we also derive expressions for multi-operator network sharing. The obtained expressions reveal that multi-operator network sharing, particularly when coupled with multi-connectivity, eases the amount of bandwidth required for ultra-reliable communication. Numerical evaluation shows that multi-operator connectivity sharing can reduce the bandwidth required by orders of magnitude, even in sharing arrangements with only two operators.

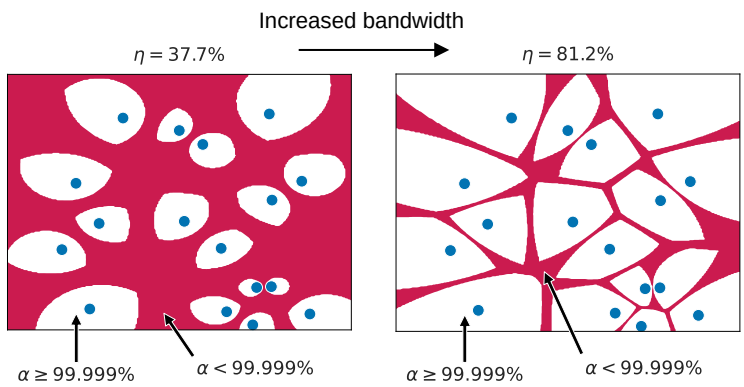


Figure 5.1: Example of a network that supports ultra-reliable communication at the link- and spatial-reliability levels α and η. Blue markers represent base stations.

5.1.1 Related work

Reliability, as defined by the 3GPP [31], is the success probability of delivering δ bits within a user-plane deadline τ, at a certain channel quality. Ultra-reliability in this context refers to the multiple "nines" that describe the stringent requirement on this probability, e.g., "five-nines" is 99.999%. Using this definition, [12] estimated the bandwidth required to support ultra-reliable communication for given signal-to-noise ratio values. While relevant for scheduling or rate control mechanisms, this approach does not translate into network-

wide spectrum requirements, for the signal-to-noise ratio or, more generally, channel quality typically varies widely throughout the network. In this regard, [6] and the work in Chapter 2 considered the spectrum needs to support ultra-reliable communication in industrial and large-scale urban networks, respectively. These works performed extensive system-level simulations to obtain the signal-to-interference-plus-noise ratio throughout the network and then map it onto the bandwidth required to support the performance targets of given ultra-reliable communication services. This approach, however, is costly because of the massive number of simulation cycles (or data samples) it takes to capture rare network conditions when the desired reliability is $\geq 99.999\%$, as we have shown in Chapter 4.

The objective of this chapter is to address the limitations of the previous studies mentioned above by developing an analytical framework to calculate the spectrum needs for ultra-reliable communication in large-scale networks. The proposed framework is based on meta distributions of the SIR introduced in [61] to study the spatial distribution of outages and later coined in [18, 22, 62] to study the spatial distribution of the SIR throughout the network. Compared to the existing literature on meta distributions (e.g., spatial distribution of the SIR [18, 22, 61, 62], power control [63], rate control [64], and age of information [65]), this chapter presents the first set of works to apply meta distributions to spectrum dimensioning in ultra-reliable communication. Furthermore, by incorporating the notion of spatial reliability, the proposed framework allows the direct assessment of the spatial reach of ultra-reliable communication in the network, which is not present in [6, 12] and previous chapters of this dissertation.

Inspired by Chapters 2 and 3, which report network sharing as an attractive way of leveraging additional network resources for enhanced reliability, this chapter also considers multi-operator network sharing, resulting in the first set of works to provide closed-form expressions for the spectrum needs for network-wide ultra-reliable communication in multi-operator net-

work sharing scenarios.

5.2 Proposed framework for spectrum dimensioning

The proposed framework differentiates between two notions of reliability and uses them to derive the required bandwidth to support ultra-reliable communication network-wide.

Definition 5.1 (Link reliability)**.** The link reliability α is defined as the success rate of transmitting δ bits within a user-plane deadline τ for a given point pattern ϕ:

$$\alpha = \Pr\left(t_\delta(w_\phi, \mathrm{SIR}) \leq \tau | \phi\right), \tag{5.1}$$

where $t_\delta(w_\phi, \mathrm{SIR})$ is a function that maps bandwidth w_ϕ and SIR to user-plane latency, and ϕ a point pattern that describes the locations of the base stations in the network.

Definition 5.2 (Spatial reliability)**.** The spatial reliability η is defined as the service area, in percentage, from which users meet the link reliability α:

$$\eta = \Pr\left(\Pr\left(t_\delta(w, \mathrm{SIR}) \leq \tau | \Phi\right) \geq \alpha\right), \tag{5.2}$$

for any ergodic point process Φ[1].

Spatial reliability provides information about the reach of ultra-reliable communication. For instance, ultra-reliable communication is supported in $\eta \approx 37\%$ (left-hand side) and $\eta \approx 81\%$ (right-hand side) of the locations in the networks in Figure 5.1. Notice that spatial reliability

[1]Notation: Random variables are denoted in upper case (e.g., Φ), whereas realizations of random variables are in lower case (e.g., $\phi \in \Phi$).

stems from a meta distribution, i.e., the distribution function of a conditional distribution [62].

We are interested in the amount of spectrum it takes to meet the link- and spatial-reliability targets α and η. To that end, Equation (5.1) and Equation (5.2) need to be inverted so that the bandwidth becomes a function of the parameters α and η. To start with, as in Chapter 2, the user-plane latency and its inverse are assumed to be functions of the Shannon-Hartley capacity,

$$\begin{aligned} t_\delta(w_\phi, \mathrm{SIR}) &= \frac{\delta}{w_\phi \log_2(1+\mathrm{SIR})} \text{ and} \\ t_\delta^{-1}(\tau, \mathrm{SIR}) &= \frac{\delta}{\tau \log_2(1+\mathrm{SIR})}. \end{aligned} \tag{5.3}$$

Being $t_\delta(w_\phi, \mathrm{SIR})$ monotonically decreasing in w_ϕ and SIR, we can also express the link reliability in terms of $t_\delta^{-1}(\tau, \mathrm{SIR})$ based on the dual property [18, Theorem 1]:

$$\Pr\left(t_\delta(w_\phi, \mathrm{SIR}) \leq \tau | \phi\right) = \Pr\left(t_\delta^{-1}(\tau, \mathrm{SIR}) \leq w_\phi \middle| \phi\right). \tag{5.4}$$

Notice that the property above is not limited to our definition in Equation (5.3) and holds for any $t_\delta(w_\phi, \mathrm{SIR})$ monotonically decreasing in w_ϕ and SIR.

The right-hand side of Equation (5.4) implies that the bandwidth required to support ultra-reliable communication at an arbitrary location can be expressed as the inverse of the CDF of $t_\delta^{-1}(\tau, \mathrm{SIR})$ conditioned to ϕ,

$$w_\phi = \mathrm{F}^{-1}_{t_\delta^{-1}(\tau,\mathrm{SIR})|\phi}(\alpha). \tag{5.5}$$

If the network is defined as a collection of point patterns according to a point process Φ,

conditioning the equation above to Φ provides us with the demand for bandwidth throughout the network, a random variable defined as:

$$W_\Phi = \mathrm{F}^{-1}_{t_\delta^{-1}(\tau,\mathrm{SIR})|\Phi}(\alpha). \tag{5.6}$$

This provides an alternative way to define spatial reliability based on the duality in Equation (5.4). Here, spatial reliability can be interpreted as the service area where links demand less or the same amount of bandwidth than what is available to the network,

$$\eta = \Pr(W_\Phi \leq w), \tag{5.7}$$

which can be inverted to obtain the minimum bandwidth required to support the reliability targets α and η,

$$w(\alpha,\eta) = \mathrm{F}^{-1}_{W_\Phi}(\eta) = \mathrm{F}^{-1}_{\mathrm{F}^{-1}_{t_\delta^{-1}(\tau,\mathrm{SIR})|\Phi}(\alpha)}(\eta). \tag{5.8}$$

As spatial reliability stems from a meta distribution, the required bandwidth to meet the reliability targets α and η is expressed in terms of the inverse of a meta distribution, i.e., the inverse distribution function of the inverse of a conditional distribution. In principle, this framework allows spectrum dimensioning for a broad range of stochastic network models. Despite its complex form, the following Sections show that the expression in Equation (5.8) can be translated into simple closed-form expressions for several Poisson network scenarios, including multi-operator network sharing.

5.3 Spectrum dimensioning in Poisson networks

This Section presents the expressions for the required bandwidth to meet the ultra-reliability target requirements α and η in the downlink in networks in which the spatial deployment of base stations is modeled according to a homogeneous Poisson point process (PPP) $\Phi \in \mathbb{R}^2$ of density λ base stations per km^2. All base stations are assumed to transmit at the same power level and frequency band (i.e., frequency reuse factor 1) and are equipped with omnidirectional antennas. Mobiles are also equipped with omnidirectional antennas and associate with the base station that provides the highest ratio of the average received power to the average interference, an association policy akin to reference signal received quality (RSRQ)-based association in modern mobile networks. Fading is Rayleigh, and path loss follows a power law expression with exponent $b > 2$. Base stations are assumed to be constantly transmitting, i.e., interference-limited networks, and interference is caused by non-serving base stations. For each realization $\phi \in \Phi$, we consider the typical user located at the origin.

5.3.1 Single-operator network

Single interferer

Let us first consider a network in which interference is limited to the nearest interferer, a highly tractable scenario for meta distributions [18] that provides a lower bound for the bandwidth required to support the reliability target parameters α and η in interference-limited networks.

Lemma 5.3. *The minimum bandwidth required to support the reliability targets* $\alpha \in [0, 1]$

and $\eta \in [0,1]$ in Poisson networks with a single interferer is (proof in Appendix A):

$$w(\alpha,\eta) = \frac{\delta}{\tau \log_2\left(\frac{1-\alpha}{\alpha\eta^{b/2}} + 1\right)}, \tag{5.9}$$

where δ (amount of data to be transmitted, in bits) and τ (user-plane deadline, in seconds) are service requirements of ultra-reliable communication services and b is the path-loss exponent.

Remark 5.4. As $\alpha \to 1$, the logarithm in Equation (5.9) can be well approximated by the first term of its Taylor series expansion:

$$w(\alpha,\eta) \approx \frac{\delta \ln 2}{\tau} \times \frac{1}{1-\alpha} \times \eta^{b/2}. \tag{5.10}$$

Equation (5.10) reveals a surprisingly straightforward relationship between spectrum needs and the reliability targets α and η. As in the meta distribution of the SIR in the downlink of Poisson networks (e.g., [18, § III.A]), the density of the PPP does not impact the bandwidth requirement. More strikingly, the link reliability requires approximately one order of magnitude increase in bandwidth per "nine", exposing a major challenge in supporting ultra-reliable communication services in wireless networks. As we would intuitively expect, the demand for bandwidth can be lowered by decreasing the reach of ultra-reliable communication in the network, reducing it to zero in the asymptotic case of $\eta \to 0$. This reinforces the intuition that achieving ultra-reliable communication may be feasible in some parts of the network, but realizing it almost everywhere can be challenging (i.e., $\alpha \to 1$ and $\eta \to 1$ simultaneously).

To quantify the relationship between spectrum and reliability targets α and η, Figure 5.2 numerically evaluates what it takes to support ultra-reliable communication throughout the network considering the reference values $\delta = 256$ bits, $\tau = 1$ ms [31], and $b = 3.5$ (path-loss exponent). Monte Carlo simulations are used to validate *Lemma* 5.3 and *Remark* 5.4. Notice that the simulations corroborate the accuracy of Equation (5.9), and the accuracy of the approximation in Equation (5.10) increases with α, as expected.

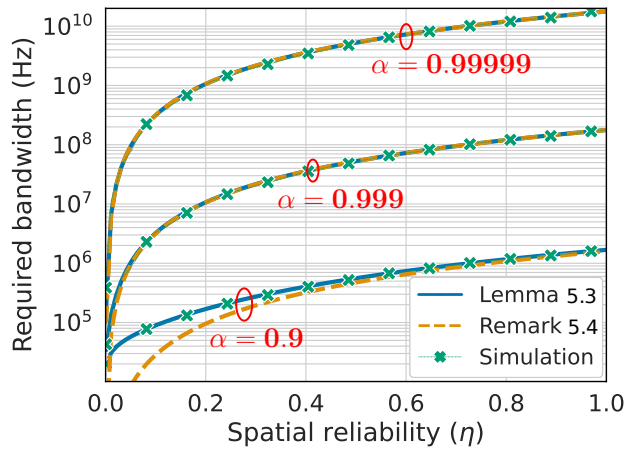


Figure 5.2: The bandwidth required to support $\alpha \in \{0.9, 0.999, 0.99999\}$ and $\eta \in [0, 1]$ when only the nearest interferer is considered for $\delta = 256$ bits, $\tau = 1$ ms, and $b = 3.5$.

Full interference

Let us now consider a network in which all interferers are taken into account, providing an upper bound for the bandwidth required to support ultra-reliable communication in interference-limited networks.

Lemma 5.5. *The minimum bandwidth required to support the reliability targets $\alpha \in [0, 1]$ and $\eta \in [0, 1]$ in Poisson networks with full interference can be approximated as (proof in Appendix B):*

$$w(\alpha, \eta) \approx \frac{\delta}{\tau \log_2 \left(1 + \left(\frac{1-\alpha}{\alpha \eta^{b/2}}\right) \times \left(\frac{b-2}{b+2}\right)\right)}. \tag{5.11}$$

(A similar approximation can be obtained following the meta distribution of the SIR in [22, Corollary 5]; however, [22] does not consider the relationship between spectrum needs and reliability targets, which is even further exposed in the *Remark* below).

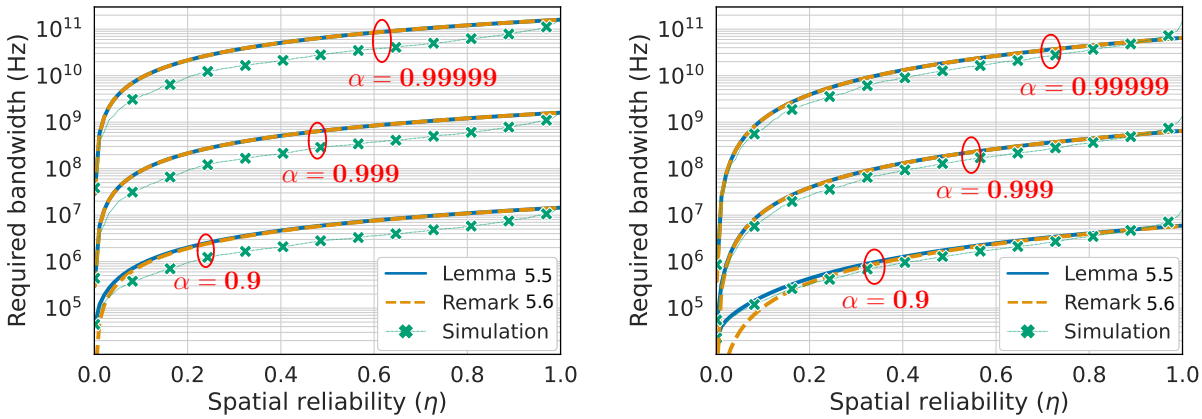
Remark 5.6. As $\alpha \to 1$, the logarithm in Equation (5.11) can be well approximated by the first term of its Taylor series expansion:

$$w(\alpha, \eta) \approx \frac{\delta \ln 2}{\tau} \times \frac{b+2}{b-2} \times \frac{1}{1-\alpha} \times \eta^{b/2}. \tag{5.12}$$

Similarly to the single-interferer scenario, Equation (5.12) reveals a surprisingly straightforward relationship between spectrum needs and reliability targets α and η. The link reliability requires approximately one order of magnitude of increase in bandwidth per "nine". The demand for bandwidth decreases as a function of the reach of ultra-reliable communication in the network, reducing it to zero in the asymptotic case of $\eta \to 0$.

Figure 5.3 is Figure 5.2's counterpart and numerically evaluates the required bandwidth when full interference is considered. Notice that our derivations approximate well the required bandwidth in the two networks assessed in Figure 5.3, particularly as the path-loss exponent increases, as we expect based on the approximations in Appendix B. As revealed in Equation (5.12), each "nine" in the link reliability translates into an order of magnitude increase in bandwidth.

In practice, interference-limited networks may operate between single-interferer and full-interference scenarios because of, for instance, line-of-sight and non-line-of-sight conditions,



(a) Path-loss exponent $b = 2.5$.

(b) Path-loss exponent $b = 3.5$.

Figure 5.3: The bandwidth required to support $\alpha \in \{0.9, 0.999, 0.99999\}$ and $\eta \in [0, 1]$ when all interferers are considered for $\delta = 256$ bits, $\tau = 1$ ms, and $b \in \{2.5, 3.5\}$.

which may (or may not) weaken the impact of far interferers on the downlink. Yet, the single-interferer and full-inference scenarios are useful because they provide us with lower and upper bounds for the required bandwidth to support the reliability targets α and η in interference-limited networks. Figure 5.4 quantifies these bounds for a network with path loss $b = 3.5$ using *Remarks* 5.4 and 5.6. In ultra-reliability regimes with $\alpha = 0.99999$, the demand for bandwidth can be tens of gigahertz, beyond what is typically available in today's networks, as opposed to a few hundred megahertz when $\alpha = 0.999$. However, that does not mean that ultra-reliable communication is not supported. Instead, ultra-reliable communication is concentrated in a handful of locations. For instance, if 1 GHz of bandwidth is made available, only 10%-20% of the network supports ultra-reliable communication at $\alpha = 0.99999$ in Figure 5.4.

These results imply that higher frequency bands are better suited for ultra-reliable communication, for they are often coupled with wider bandwidth than lower frequency bands. Furthermore, our derivations suggest another attractive characteristic of higher bands for ultra-reliable communication. Higher frequency bands are often associated with greater path-loss exponents. In our derivations, the bandwidth requirement is inversely proportional to

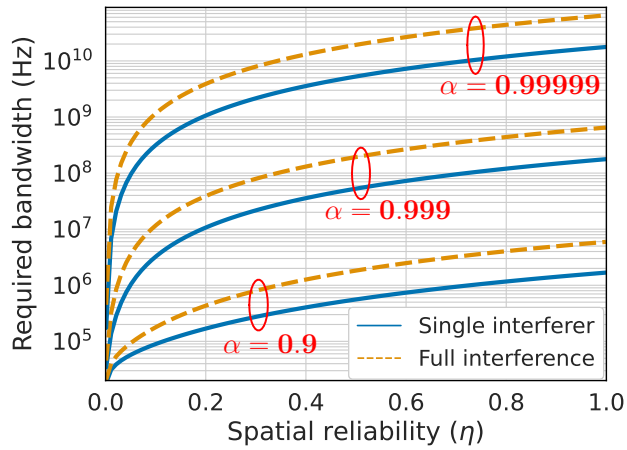


Figure 5.4: The bandwidth required to support $\alpha \in \{0.9, 0.999, 0.99999\}$ and $\eta \in [0, 1]$ in single-interferer vs. full-interference scenarios for $\delta = 256$ bits, $\tau = 1$ ms, and $b = 3.5$.

the path-loss exponent b (e.g., $\eta^{b/2} \to 0$ as $b \to \infty$ in *Remarks* 5.4 and 5.6), suggesting that higher bands may demand less bandwidth than lower bands, or, analogously, that higher bands may expand the reach of ultra-reliable communication for the same amount of bandwidth. For instance, in Figure 5.3, the reach of ultra-reliable communication at $\alpha = 0.99999$ expands from 20% in a network with path-loss exponent $b = 2.5$ (Figure 5.3a) to 40% in a network with path-loss exponent $b = 3.5$ (Figure 5.3b) for the same 10 GHz of bandwidth made available.

5.3.2 Multi-operator network sharing

Motivated by Chapters 2 and 3, this Section considers the impact of multi-operator network sharing on the spectrum needs for network-wide ultra-reliable communication. The multi-operator network consists of k operators. Base stations are deployed according to a PPP Φ_i for each operator i. Φ_i and Φ_j are assumed to be independent for all $i \neq j$. We focus on the full-interference scenario. Mobile operators operate in their respective licensed bands. This way, interference is caused by non-serving base stations within the same network operator,

with no cross-operator interference.

Multi-operator network sharing with single-connectivity

We start by considering the spectrum needs for multi-operator network sharing with single-connectivity. In this scenario, a mobile connects to a base station of either operator in the multi-operator network: the base station that provides the highest ratio of the average received power to the average interference.

Lemma 5.7. *The bandwidth required to support the reliability targets $\alpha \to 1$ and $\eta \in [0,1]$ in multi-operator network sharing with single-connectivity with k operators can be approximated as (proof in Appendix C):*

$$w(\alpha,\eta) \approx \frac{k\delta}{\tau \log_2\left(1 + \frac{1-\alpha}{\alpha} \times \left(\frac{\mathrm{sinc}(2/b)}{1-(1-\eta)^{\frac{1}{k}}}\right)^{\frac{b}{2}}\right)}, \tag{5.13}$$

where δ (amount of data to be transmitted, in bits) and τ (user-plane deadline, in seconds) are service requirements of ultra-reliable communication and b is the path-loss exponent.

Remark 5.8. As $\alpha \to 1$, we can simplify Equation (5.13) by approximating the logarithm to the first term of its Taylor series expansion:

$$w(\alpha,\eta) \approx \frac{k\delta \ln(2)}{\tau} \times \left(\frac{1-(1-\eta)^{\frac{1}{k}}}{\mathrm{sinc}(2/b)}\right)^{\frac{b}{2}} \times \frac{1}{1-\alpha}. \tag{5.14}$$

As in the single-operator network scenario, *Remark* 5.8 reveals a surprisingly simple rela-

tionship between spectrum needs and reliability requirements, where each "nine" in the link reliability α corresponds to one order of magnitude in bandwidth. Figure 5.5 quantifies this relationship in a dual-operator network sharing arrangement. Monte Carlo simulations are used to validate *Lemma* 5.7 and *Remark* 5.8. Notice that *Lemma* 5.7 captures well the relationship between spectrum needs and reliability targets. As expected, the accuracy of *Remark* 5.8 increases as $\alpha \to 1$, capturing well the spectrum needs in ultra-reliability regimes, where α is generally ≥ 0.99999.

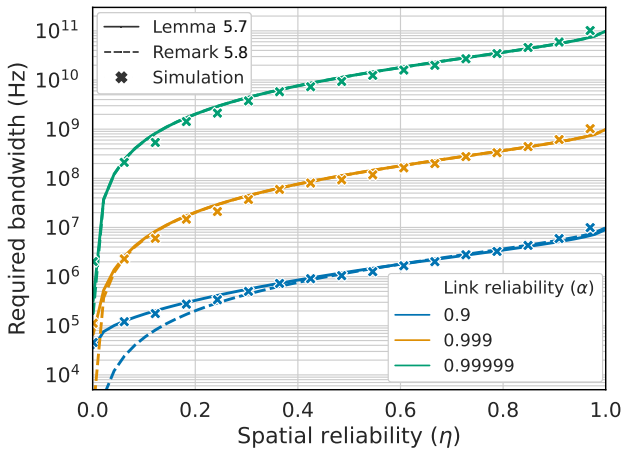


Figure 5.5: The bandwidth required to support $\alpha \in \{0.9, 0.999, 0.99999\}$ and $\eta \in [0, 1]$ in dual-operator network sharing (i.e., $k = 2$) with single-connectivity for $\delta = 256$ bits, $\tau = 1$ ms, and $b = 3.5$.

Multi-operator network sharing with multi-connectivity

In multi-operator network sharing with multi-connectivity, a.k.a. multi-operator connectivity sharing, a mobile connects to a base station of each operator in the multi-operator sharing arrangement. Signals from the different connections are combined according to selection combining, for it does not require tight synchronization compared to other techniques such as equal-gain and maximal-ratio combining while yielding similar reliability gains [66].

Lemma 5.9. *The bandwidth required to support the reliability targets* $\alpha \to 1$ *and* $\eta \in [0, 1]$

in multi-operator network sharing with multi-connectivity to k operators can be approximated as (proof in Appendix D):

$$w(\alpha,\eta) \approx \left[\frac{k\delta}{\tau \log_2\left(1 + \frac{(1-\alpha)^{1/k}}{1-(1-\alpha)^{1/k}} \times \left(\frac{\mathrm{sinc}(2/b)}{\eta^{1/k}}\right)^{\frac{b}{2}}\right)} \times \frac{k\delta}{\tau \log_2\left(1 + \frac{(1-\alpha)^{1/k}}{1-(1-\alpha)^{1/k}} \times \left(\frac{\mathrm{sinc}(2/b)}{1-(1-\eta)^{1/k}}\right)^{\frac{b}{2}}\right)} \right]^{\frac{1}{2}}. \tag{5.15}$$

The relationship between spectrum needs and reliability targets α and η is subtler in *Lemma* 5.9 than in its counterparts. Particularly, Equation (5.15) cannot be approximated with its Taylor series expansion because its derivative is undefined as $\alpha \to 1$ for $k > 1$, preventing simplifications as in *Remark* 5.8. Figure 5.6 numerically evaluates *Lemma* 5.9 for the same reference values as in Figure 5.5 in dual-operator network sharing with dual-connectivity. Again, Monte Carlo simulations are used to validate the accuracy of our derivations. The bandwidth requirement in Figure 5.6 is lower than in Figure 5.5, indicating that multi-operator network sharing benefits from multi-connectivity to further lower the demand for spectrum in ultra-reliability regimes. For instance, if 1 GHz of bandwidth is made available, Figure 5.6 shows that dual-operator network sharing with dual-connectivity can support "five-nines" link reliability (i.e., $\alpha = 0.99999$) throughout the entire network (i.e., $\eta \to 1$), whereas Figure 5.5 shows that dual-operator network sharing with single-connectivity only supports "five-nines" in 15% of the network area (i.e., $\eta = 0.15$).

While both multi-operator network sharing scenarios benefit from an increase in the number of operators in the multi-operator network, the impact of k is different on single- and multi-

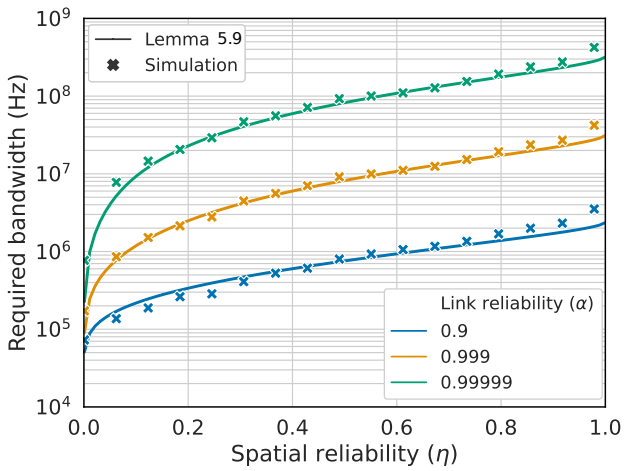


Figure 5.6: The bandwidth required to support $\alpha \in \{0.9, 0.999, 0.99999\}$ and $\eta \in [0, 1]$ in dual-operator network sharing with dual-connectivity (i.e., $k = 2$) for $\delta = 256$ bits, $\tau = 1$ ms, and $b = 3.5$.

connectivity. This is illustrated in Figure 5.7. *Lemmas* 5.7 and 5.9 yield the same expression for the special case of $k = 1$, as both scenarios correspond to single-operator networks with single-connectivity (denoted as SoP-SC). In fact, this special case is akin to the derivations in Appendix B for *Lemma* 5.5 and results in equivalent approximations, where $\text{sinc}^{\frac{b}{2}}\left(\frac{2}{b}\right)$ is replaced with $\frac{b-2}{b+2}$. The crossover points in the upper right quadrant of Figure 5.7 are due to approximation inaccuracies in Equation (5.11) when $\eta \to 1$, as shown in Figure 5.3b. As k increases, multi-operator network sharing with multi-connectivity (denoted as MoP-MC) further reduces the bandwidth requirement compared to multi-operator network sharing with single-connectivity (denoted as MoP-SC). Particularly, multi-connectivity reduces the required bandwidth by approximately two orders of magnitude in $k = 2$. While increasing k further eases the bandwidth requirement, the gains become less significant for $k > 4$. This is good news because it indicates that even dual-operator network sharing agreements, when coupled with multi-connectivity, can yield substantial gains in terms of spectrum needs for network-wide ultra-reliable communication.

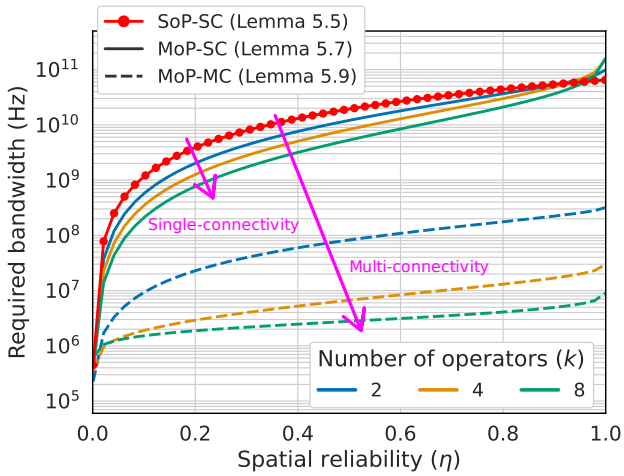


Figure 5.7: The bandwidth required to support $\alpha = 0.99999$ and $\eta \in [0, 1]$ in multi-operator network sharing with single- and multi-connectivity for $\delta = 256$ bits, $\tau = 1$ ms, and $b = 3.5$.

5.4 Chapter summary

This chapter proposed a framework based on meta distributions to dimension spectrum for network-wide ultra-reliable communication. Using the proposed framework, we obtained closed-form expressions for the required bandwidth to support the link-level reliability α in a proportion η of the network area in Poisson networks. These expressions reveal an interesting relationship between spectrum needs and reliability requirements, in which each "nine" in reliability roughly translates into an order of magnitude increase in bandwidth. In ultra-reliability regimes, where α is generally ≥ 0.99999, the required bandwidth can be in the order of tens (or even hundreds) of gigahertz, suggesting that higher frequency bands, which are generally coupled with wider bandwidth, may be attractive to support ultra-reliable communication. Furthermore, we considered multi-operator network sharing. Particularly, multi-operator connectivity sharing, which combines multi-operator network sharing with multi-connectivity, is shown to significantly reduce the demand for spectrum, being an attractive alternative to ease the amount of spectrum required for network-wide ultra-reliable communication.

Chapter 6

Conclusion and future work

6.1 Summary

In this dissertation, we studied wireless network dimensioning and provisioning for ultra-reliable communication services, an important step to enabling emerging mission-critical services such as URLLC and HRLLC in future mobile networks, for which reliable and resilient communication is a must. In particular, this dissertation focused on spectrum dimensioning and studied what it takes – e.g., suitable spectrum bands and required bandwidth – to meet ultra-reliable communication services' stringent reliability and latency requirements.

Chapter 2 quantified the required bandwidth to support ultra-reliable communication for low-, mid-, and high-frequency bands using network models by the 3GPP. In ultra-reliability regimes, the required bandwidth can be in the order of gigahertz, suggesting that high-frequency bands are an attractive alternative for ultra-reliable communication – despite their adverse propagation characteristics – because they are often coupled with wider bandwidth. Yet, provisioning bandwidth in the order of gigahertz can be challenging because it is beyond what is typically available (or even practical) in today's networks. Alternatively, we consider multi-operator network sharing. In particular, we show that multi-operator connectivity sharing, which combines multi-operator network sharing and multi-connectivity, can significantly reduce the spectrum needs for ultra-reliable communication, lowering the

required bandwidth by a few orders of magnitude in ultra-reliability regimes.

Chapter 3 further studied multi-operator connectivity sharing for ultra-reliable communication. Using a real-world dataset containing signal-strength indicators collected in the city of Dublin, Ireland, for three mobile network operators, Chapter 3 shows that multi-operator connectivity sharing is attractive because mobile operators tend to exhibit complementary characteristics that benefit reliable communication, such as complementary coverage during periods of impaired connectivity. However, multi-operator connectivity sharing comes at the expense of increased network load because of the additional connections required, which, in turn, can throttle gains in multi-operator connectivity sharing when all mobiles in the network multi-connect. A simple but effective alternative is to only allow mobiles with weak signal strength to multi-connect, which is shown to mitigate the drawbacks of multi-operator connectivity sharing while achieving increased reliability.

Chapter 4 addressed the practical challenge of assessing ultra-reliability regimes for resource dimensioning for ultra-reliable wireless communication. As we have shown, it can take massive amounts of data to assess rare network conditions that may only occur $\leq 0.001\%$ of the time when reliability is $\geq 99.999\%$, which can be time-consuming (or even impractical) depending on the time it takes to collect each data sample. To circumvent this, we proposed an extreme value theory (EVT)-based method for spectrum dimensioning. The proposed method reduces the amount of data samples required for spectrum dimensioning in ultra-reliability regimes by several orders of magnitude compared to what would be needed using traditional or alternative state-of-the-art methods.

Lastly, Chapter 5 provided analytical models for spectrum dimensioning for ultra-reliable communication. We proposed a framework, based on meta distributions, that differentiates between two notions of reliability: (a) the link reliability α, which corresponds to the suc-

cess rate of meeting the performance-level requirements of URLLC and HRLLC services, and (b) the spatial reliability η, which corresponds to the proportion of the service area in which these services are supported. Using our framework, we obtained closed-form expressions for the spectrum needs to support the reliability target requirements α and η for typical single-operator network scenarios as well as multi-operator network sharing scenarios. These expressions reveal an interesting relationship between spectrum needs and reliability targets – in which each "nine" in reliability can translate into an order of magnitude increase in bandwidth – and how multi-operator connectivity sharing can reduce the demand for spectrum in ultra-reliability regimes.

6.1.1 Potential future directions

Multi-operator connectivity sharing. The works in this dissertation reveal that multi-operator connectivity sharing may be an attractive way of enabling ultra-reliable communication in mobile networks. Multi-operator network sharing is akin to today's mobile virtual network operators that operate on top of a multi-operator network, such as Google Fi, which formerly operated on top of a multi-carrier network. Yet, in today's multi-operator networks, mobiles are often subject to resource management policies by underlying operators. Network slicing is a potential solution for reserving and isolating resources. However, coordinating resources from independently managed underlying networks can be particularly challenging. A potential future research direction is to investigate the challenges involved in achieving network slicing in multi-operator networks for ultra-reliable communication, complementing efforts by the 3GPP in TR 28.811 to standardize multi-operator network slicing. Furthermore, as pointed out in Chapter 3, multi-operator connectivity sharing incurs increased network load, which, in turn, can jeopardize reliability. We investigated a simple heuristic

in Chapter 3 to circumvent this drawback. Yet, multi-operator connectivity sharing may further benefit from finer-grained methods to cope with increased network load, increasing network reliability while reducing the impact of increased load.

End-to-end reliability. The works in this dissertation focus on achieving ultra-reliable communication considering the communication between mobiles and base stations, being agnostic to network components beyond the cell site. However, reliability is ultimately end-to-end. As illustrated in Figure 6.1, a potential research direction is to consider the network beyond the cell site, including fronthaul, midhaul, backhaul, and core network.

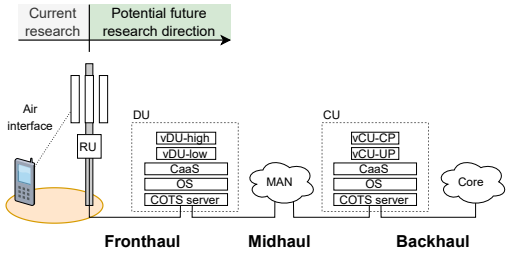


Figure 6.1: Open RAN pipeline example.

Reconfigurable intelligent surfaces (RIS) to support ultra-reliable communication. Another potential research direction is to consider emerging technologies such as reconfigurable intelligent surfaces to support ultra-reliable communication. Reconfigurable intelligent surfaces are programmable passive relays that can enhance communication, particularly in adverse channel conditions, as illustrated in Figure 6.2. They may provide an alternative means of supporting ultra-reliable communication in large-scale mobile networks. In light of the work in this dissertation, reconfigurable intelligent surfaces could be modeled as a network asset for deployment, as spectrum and network density in Chapter 2. This research direction aligns well with recent efforts for standardization of reconfigurable intelligent surfaces by the 3GPP and ETSI and could help to shed light on RIS deployment scenarios [67].

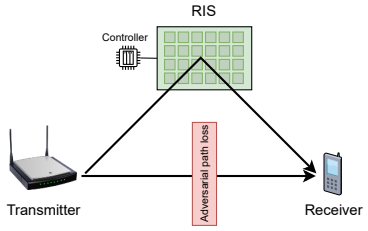


Figure 6.2: Example of RIS-aided wireless communications.

6.2 Other research activities beyond this dissertation

During my Ph.D., I have also engaged in other research activities beyond what has been listed in this dissertation.

Research collaboration. I have collaborated with the 6G Flagship initiative (www.6gflagship.com) on the topic of securing reconfigurable intelligent surface-aided wireless communications as part of our research group's collaboration with the University of Oulu, Finland. To this day, this collaboration has resulted in two manuscripts:

1. A. S. de Sena, J. Kibiłda, N. H. Mahmood, **A. Gomes**, and M. Latva-aho, "Malicious RIS versus massive MIMO: Securing multiple access against RIS-based jamming attacks," submitted to IEEE Wireless Communications Letters, 2023.

2. J. Kibiłda, N. H. Mahmood, **A. Gomes**, M. Latva-aho, L. A. DaSilva, "Reconfigurable intelligent surfaces: The next frontier of next G security," in arXiv, 2022, DOI: arxiv.org/abs/2212.05101.

Book chapter. The work on reconfigurable intelligent surfaces has also been included in a book chapter on securing millimeter-wave communication systems that is currently being considered for publication:

1. J. Kibiłda, K. Zeng, P. Pathak, N. N. Sapavath, **A. Gomes**, and V. K. Shah, "Securing next G millimeter-wave," 2023. (In progress)

Internship. I interned at AT&T Research Labs in the Summer of 2023. During my internship, I worked on modeling and assessing resiliency on Open RAN architectures, including the characterization of failures and repairs of RAN components.

Presentations. I have also presented my research during conferences and workshops for further dissemination, including:

1. "Analysis of spectrum needs for network-wide ultra-reliable communication with network sharing," in IEEE Globecom Workshops, Kuala Lumpur, Malaysia, 2023. (Upcoming)

2. "Dimensioning spectrum for ultra-reliable communication in next G networks," in Commonwealth Cyber Initiative (CCI) Symposium, Richmond, VA, 2023.

3. "Reconfigurable intelligent surfaces: The new frontier of next G security," in WISPER Center Planning Workshop, Arlington, VA, 2022.

4. "Achieving ultra-reliable communication in mobile networks," in Commonwealth Cyber Initiative (CCI) Symposium, Richmond, VA, 2022.

5. "Network sharing for reliable networks: A data-driven study," in IEEE ICC, Dublin, Ireland, 2020.

Bibliography

[1] S. Chandrashekar, A. Maeder, C. Sartori, T. Höhne, B. Vejlgaard, and D. Chandramouli, "5G multi-RAT multi-connectivity architecture," in *IEEE International Conference on Communications workshops (ICC)*, pp. 180–186, 2016.

[2] P. Popovski, J. J. Nielsen, C. Stefanovic, E. De Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park, *et al.*, "Wireless access for ultra-reliable low-latency communication: principles and building blocks," *IEEE Transactions on Communications*, vol. 32, no. 2, pp. 16–23, 2018.

[3] Next G Alliance, "Next G alliance report: 6G applications and use cases," tech. rep., Next G Alliance, 5 2022.

[4] ITU-R, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," Tech. Rep. ITU-R M.2410-0, International Telecommunication Union (ITU), 11 2017.

[5] ITU, "IMT towards 2030 and beyond." https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2030/Pages/default.aspx. Accessed: Sept 2023.

[6] O. N. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *2015 IEEE international conference on communication workshop (ICCW)*, pp. 1190–1195, IEEE, 2015.

[7] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang,

M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2488–2524, 2019.

[8] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. Santos Filho, and G. Fettweis, "How reliable and capable is multi-connectivity?," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1506–1520, 2018.

[9] J. J. Nielsen, R. Liu, and P. Popovski, "Ultra-reliable low latency communication using interface diversity," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1322–1334, 2018.

[10] A. Anand, G. De Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 477–490, 2020.

[11] "The twelve networking truths." RFC 1925, Apr. 1996.

[12] P. Popovski, . Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, 2019.

[13] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[14] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of banking & finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[15] R. Tyrrell Rockafellar and J. O. Royset, "Engineering decisions under risk averseness," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, vol. 1, no. 2, p. 04015003, 2015.

[16] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*, vol. 208. Springer, 2001.

[17] M. Haenggi, "Meta distributions—part 1: Definition and examples," *IEEE Communications Letters*, vol. 25, no. 7, pp. 2089–2093, 2021.

[18] M. Haenggi, "Meta distributions—part 2: Properties and interpretations," *IEEE Communications Letters*, vol. 25, no. 7, pp. 2094–2098, 2021.

[19] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Communications Letters*, vol. 23, no. 4, pp. 740–743, 2019.

[20] C.-F. Liu, M. Bennis, and H. V. Poor, "Latency and reliability-aware task offloading and resource allocation for mobile edge computing," in *2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–7, 2017.

[21] N. Mehrnia and S. Coleri, "Wireless channel modeling based on extreme value theory for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1064–1076, 2022.

[22] K. Feng and M. Haenggi, "Separability, asymptotics, and applications of the SIR meta distribution in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4806–4816, 2020.

[23] "Transmission Control Protocol." RFC 793, Sept. 1981.

[24] M. Haenggi and R. Smarandache, "Diversity polynomials for the analysis of temporal correlations in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5940–5951, 2013.

[25] M.-T. Suer, C. Thein, H. Tchouankem, and L. Wolf, "Multi-connectivity as an enabler for reliable low latency communications—an overview," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 156–169, 2020.

[26] M. Oikonomakou, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Cooperative base station switching off in multi-operator shared heterogeneous network," in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2015.

[27] P. Di Francesco, F. Malandrino, T. K. Forde, and L. A. DaSilva, "A sharing- and competition-aware framework for cellular network evolution planning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 2, pp. 230–243, 2015.

[28] J. Kibiłda, N. J. Kaminski, and L. A. DaSilva, "Radio access network and spectrum sharing in mobile networks: A stochastic geometry perspective," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2562–2575, 2017.

[29] P. Di Francesco, F. Malandrino, and L. A. DaSilva, "Mobile network sharing between operators: a demand trace-driven study," in *Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing*, pp. 39–44, 2014.

[30] Next G Alliance, "Next G alliance report: roadmap to 6G," tech. rep., Next G Alliance, 2 2022.

[31] 3GPP, "Study on scenarios and requirements for next generation access technologies," Tech. Rep. 38.913, 3rd Generation Partnership Project (3GPP), 07 2020. Version 16.0.0.

[32] 3GPP, "Technical specification group radio access network; study on channel model for frequencies from 0.5 to 100 GHz," Tech. Rep. 38.901, 3rd Generation Partnership Project (3GPP), 12 2019. Version 16.1.0.

[33] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LOS and NLOS transmissions in dense cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2365–2380, 2016.

[34] M. Rebato, M. Mezzavilla, S. Rangan, F. Boccardi, and M. Zorzi, "Understanding noise and interference regimes in 5G millimeter-wave cellular networks," in *European Wireless 2016; 22th European Wireless Conference*, pp. 1–5, 2016.

[35] J. Kibiłda, B. Galkin, and L. A. DaSilva, "Modelling multi-operator base station deployment patterns in cellular networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 3087–3099, 2016.

[36] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[37] A. Frommgen, T. Erbshäußer, A. Buchmann, T. Zimmermann, and K. Wehrle, "ReMP TCP: Low latency multipath TCP," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–7, 2016.

[38] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "TCP extensions for multipath operation with multiple addresses," RFC 6824, IETF, January 2013. http://www.rfc-editor.org/rfc/rfc6824.txt.

[39] M. Simsek, T. Hößler, E. Jorswieck, H. Klessig, and G. Fettweis, "Multiconnectivity in multicellular, multiuser systems: A matching- based approach," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 394–413, 2019.

[40] P. Rost, M. Breitbach, H. Roreger, B. Erman, C. Mannweiler, R. Miller, and I. Viering, "Customized industrial networks: Network slicing trial at Hamburg seaport," *IEEE Wireless Communications*, vol. 25, no. 5, pp. 48–55, 2018.

[41] A. H. Mahdi, T. Hößler, N. Franchi, and G. Fettweis, "Multi-connectivity for reliable wireless industrial communications: gains and limitations," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–7, 2020.

[42] N. Schwarzenberg, A. Wolf, N. Franchi, and G. Fettweis, "Quantifying the gain of multi-connectivity in wireless LAN," in *2018 European Conference on Networks and Communications (EuCNC)*, pp. 16–20, 2018.

[43] J. Rao and S. Vrzic, "Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis," *IEEE Network*, vol. 32, no. 2, pp. 32–40, 2018.

[44] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile network resource sharing options: Performance comparisons," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4470–4482, 2013.

[45] M. Andrews, M. Bradonjić, and I. Saniee, "Quantifying the benefits of infrastructure sharing," in *Proceedings of the 12th Workshop on the Economics of Networks, Systems and Computation*, NetEcon '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[46] Knuetter, C., "G-MoN." https://play.google.com/store/apps/details?id=de.carknue.gmon2. Accessed: Oct 2019.

[47] "The world's largest open database of cell towers." https://opencellid.org/. Accessed: Jan 2020.

[48] G. L. Horowitz, "Reference intervals: Practical aspects," *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*, vol. 19, no. 2, p. 95, 2008.

[49] P. C. F. Eggers, M. Angjelichinoski, and P. Popovski, "Wireless channel modeling perspectives for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2229–2243, 2019.

[50] A. Brighente, J. Mohammadi, P. Baracca, S. Mandelli, and S. Tomasin, "Interference prediction for low-complexity link adaptation in beyond 5G ultra-reliable low-latency communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[51] W. Zhang, M. Derakhshani, and S. Lambotharan, "Non-parametric statistical learning for URLLC transmission rate control," in *IEEE International Conference on Communications (ICC)*, pp. 1–6, 2021.

[52] M. Angjelichinoski, K. F. Trillingsgaard, and P. Popovski, "A statistical learning approach to ultra-reliable low latency communication," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5153–5166, 2019.

[53] N. Mehrnia and S. Coleri, "Extreme value theory based rate selection for ultra-reliable communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6727–6731, 2022.

[54] R. J. Hyndman and Y. Fan, "Sample quantiles in statistical packages," *The American Statistician*, vol. 50, no. 4, pp. 361–365, 1996.

[55] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: concepts, techniques and tools*. Princeton university press, 2005.

[56] S. D. Grimshaw, "Computing maximum likelihood estimates for the generalized Pareto distribution," *Technometrics*, vol. 35, no. 2, pp. 185–191, 1993.

[57] C. Scarrott and A. MacDonald, "A review of extreme value threshold estimation and uncertainty quantification," *REVSTAT–Statistical Journal*, vol. 10, no. 1, pp. 33–60, 2012.

[58] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a "kneedle" in a haystack: Detecting knee points in system behavior," in *2011 31st international conference on distributed computing systems workshops*, pp. 166–171, IEEE, 2011.

[59] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[60] "scipy.stats.gaussian_kde – SciPy v1.11.2 manual." https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html#scipy.stats.gaussian_kde. Accessed: Sept 2023.

[61] R. K. Ganti and J. G. Andrews, "Correlation of link outages in low-mobility spatial wireless networks," in *IEEE Asilomar*, pp. 312–316, 2010.

[62] M. Haenggi, "The meta distribution of the SIR in Poisson bipolar and cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2577–2589, 2015.

[63] Y. Wang, M. Haenggi, and Z. Tan, "The meta distribution of the SIR for cellular networks with power control," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1745–1757, 2017.

[64] S. S. Kalamkar and M. Haenggi, "Per-link reliability and rate control: Two facets of the SIR meta distribution," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1244–1247, 2019.

[65] P. D. Mankar, M. A. Abd-Elmagid, and H. S. Dhillon, "Spatial distribution of the mean peak age of information in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4465–4479, 2021.

[66] A. Gomes, J. Kibiłda, A. Farhang, R. Farrell, and L. A. DaSilva, "Network sharing for reliable networks: A data-driven study," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[67] ETSI, "Reconfigurable intelligent surfaces (RIS); use cases, deployment scenarios and requirements." https://www.etsi.org/deliver/etsi_gr/RIS/001_099/001/01.01.01_60/gr_RIS001v010101p.pdf. v.1.1.1, 2023. Accessed: 09/04/2023.

[68] M. Haenggi, "The mean interference-to-signal ratio and its key role in cellular and amorphous networks," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 597–600, 2014.

Appendices

Appendix A

Proof of Lemma 5.3

This proof follows the derivation of the meta distribution of the SIR in [18, § III.A]. In the single interferer model, the SIR can be expressed as:

$$\mathrm{SIR} = \frac{H_1 R_1^{-b}}{H_2 R_2^{-b}}, \tag{A.1}$$

where H_i and R_i are the fading and distance to the ith base station. Index $i = 1$ represents the serving base station and $i = 2$ the nearest interferer.

Using the definitions in Equations 5.1, 5.3, and 5.4, the link reliability for a realization ϕ of the PPP Φ can be expressed as:

$$\begin{aligned}
\alpha &= \Pr(t^{-1}(\tau, \mathrm{SIR}) \leq w_\phi | \phi) \\
&= \Pr\left(\frac{\delta}{\tau \log_2 \left(1 + \frac{H_1 r_1^{-b}}{H_2 r_2^{-b}}\right)} \leq w_\phi \middle| \phi \right) \\
&= \Pr\left(\frac{H_1}{H_2} \geq \left(2^{\frac{\delta}{\tau w_\phi}} - 1\right) \times \left(\frac{r_1}{r_2}\right)^{-b} \middle| \phi \right) \\
&\overset{(a)}{=} \frac{1}{1 + \left(2^{\frac{\delta}{\tau w_\phi}} - 1\right) \times \left(\frac{r_1}{r_2}\right)^{b}}.
\end{aligned} \tag{A.2}$$

Notice that, when conditioned to a realization ϕ, the distances to serving base station and interferer are known, meaning that r_1 and r_2 are scalars. Step (a) follows from the observation

that $\overline{\mathrm{F}}_{H_1/H_2}(x) = \frac{1}{1+x}$ [18], where $\overline{\mathrm{F}}$ is the complementary CDF, for i.i.d. $H_j \sim \mathrm{Exp}(1)$ (Rayleigh fading).

This can be rearranged with w_ϕ on the left-hand side, resulting in

$$w_\phi = \frac{\delta}{\tau \log_2\left(\left(\frac{r_1}{r_2}\right)^{-b} \times \frac{1-\alpha}{\alpha} + 1\right)}. \tag{A.3}$$

Following Equation (5.6), the random variable W_Φ can be obtained by conditioning w_ϕ to the point process Φ, meaning that R_1 and R_2 are now random variables,

$$W_\Phi = \frac{\delta}{\tau \log_2\left(\left(\frac{R_1}{R_2}\right)^{-b} \times \frac{1-\alpha}{\alpha} + 1\right)}. \tag{A.4}$$

This way, using Equation (5.7), the spatial reliability can be expressed as:

$$\begin{aligned}
\eta &= \Pr(W_\Phi \leq w) \\
&= \Pr\left(\frac{\delta}{\tau \log_2\left(\left(\frac{R_1}{R_2}\right)^{-b} \times \frac{1-\alpha}{\alpha} + 1\right)} \leq w\right) \\
&= \Pr\left(\frac{R_1}{R_2} \leq \left[\frac{1-\alpha}{\alpha\left(2^{\frac{\delta}{\tau w}} - 1\right)}\right]^{\frac{1}{b}}\right) \\
&\overset{(a)}{=} \left[\frac{1-\alpha}{\alpha\left(2^{\frac{\delta}{\tau w}} - 1\right)}\right]^{\frac{2}{b}},
\end{aligned} \tag{A.5}$$

where (a) follows that $\mathrm{F}_{R_1/R_2}(x) = x^2$ for a PPP over $\mathbb{R}^2$ [22]. The minimum bandwidth required to support the reliability target requirements α and η can be obtained by rearranging

the equation above with w on the left-hand side,

$$w(\alpha, \eta) = \frac{\delta}{\tau \log_2 \left(\frac{1-\alpha}{\alpha \eta^{b/2}} + 1 \right)}, \tag{A.6}$$

as in Equation (5.9).

Appendix B

Proof of Lemma 5.5

In the full interference scenario, the SIR can be expressed as follows:

$$\mathrm{SIR} = \frac{H_1 R_1^{-b}}{\sum_{i=2}^{\infty} H_i R_i^{-b}}, \tag{B.1}$$

where H_i and R_i are the fading and distance to the ith base station ($i = 1$ represents the serving base station; $i \geq 2$ represents the ith nearest interferer).

Using the definitions in Equation (5.1), Equation (5.3), and Equation (5.4), the link reliability for a realization ϕ of the PPP Φ can be expressed as:

$$\begin{aligned}
\alpha &= \Pr(t^{-1}(\tau, \mathrm{SIR}) \leq w_\phi | \phi) \\
&\overset{(a)}{=} \Pr\left(\frac{H_1 r_1^{-b}}{\sum_{i=2}^{\infty} H_i r_i^{-b}} \geq 2^{\frac{\delta}{\tau w_\phi}} - 1 \middle| \phi \right) \\
&\overset{(b)}{=} \Pr\left(\frac{H_1}{H_2} \geq (2^{\frac{\delta}{\tau w_\phi}} - 1) \left(t_2^b + \sum_{i=3}^{\infty} Y_i t_i^b \right) \middle| \phi \right) \\
&\overset{(c)}{\approx} \Pr\left(\frac{H_1}{H_2} \geq (2^{\frac{\delta}{\tau w_\phi}} - 1) \left(t_2^b + \sum_{i=3}^{\infty} t_i^b \right) \middle| \phi, \sum_{i=3}^{\infty} Y_i t_i^b = \mathbb{E}\left[\sum_{i=3}^{\infty} Y_i t_i^b \right] \right) \\
&\overset{(d)}{=} \frac{r_1^{-b}}{r_1^{-b} + \sum_{i=2}^{\infty} r_i^{-b} (2^{\frac{\delta}{\tau w_\phi}} - 1)},
\end{aligned} \tag{B.2}$$

where $t_i = r_1/r_i$ and $Y_i = H_i/H_2$. The expression in (a) cannot be easily dealt with. We rearrange the complementary CDF above in terms of H_1 and H_2, as in (b). In the

absence of a closed-form solution for the summation in (b), we denote it as an expectation, approximating the complementary CDF in (b) to the conditional complementary CDF in (c). This is a reasonable approximation because $t_{i\geq 3}^{b} \to 0$ for $t_i \in [0,1]$ and $b > 2$, meaning that the complementary CDF in (b) approaches the complementary CDF of the ratio of H_1 to H_2, particularly for large path-loss exponents. For i.i.d $H_i \sim \text{Exp}(1)$ (Rayleigh fading), the expectation $\mathbb{E}[Y_i]$ is undefined; however, it can be approximated by the lower bound of Jensen's inequality, which follows that $\mathbb{E}[Y_i] \geq 1$. Step (d) follows that $\overline{\text{F}}_{H_1/H_2}(x) = \frac{1}{1+x}$ for i.i.d $H_i \sim \text{Exp}(1)$ [18].

The expression above can be rearranged with w_ϕ on the left-hand side,

$$w_\phi = \frac{\delta}{\tau \log_2\left(\frac{r_1^{-b}}{\sum_{i=2}^{\infty} r_i^{-b}} \times \frac{1-\alpha}{\alpha} + 1\right)}. \tag{B.3}$$

Following Equation (5.6), the random variable W_Φ can be obtained by conditioning w_ϕ to the point process Φ, meaning that R_1 and R_i are now random variables.

$$W_\Phi = \frac{\delta}{\tau \log_2\left(\frac{R_1^{-b}}{\sum_{i=2}^{\infty} R_i^{-b}} \times \frac{1-\alpha}{\alpha} + 1\right)}. \tag{B.4}$$

This way, using Equation (5.7), the spatial reliability can be expressed as:

$$
\begin{aligned}
\eta &= \Pr\left(W_\Phi \leq w\right) \\
&\overset{(a)}{=} \Pr\left(\frac{R_1^{-b}}{\sum_{i=2}^{\infty} R_i^{-b}} \geq \frac{\alpha(2^{\frac{\delta}{\tau w}}-1)}{1-\alpha}\right) \\
&\overset{(b)}{=} \Pr\left(\frac{R_1}{R_2} \leq \left[\frac{1-\alpha}{\alpha(2^{\frac{\delta}{\tau w}}-1)\sum_{i=2}^{\infty}(R_2/R_i)^b}\right]^{\frac{1}{b}}\right) \\
&\overset{(c)}{\approx} \Pr\left(\frac{R_1}{R_2} \leq \left[\frac{(1-\alpha)(b-2)}{\alpha(2^{\frac{\delta}{\tau w}}-1)(b+2)}\right]^{\frac{1}{b}} \middle| \sum_{i=2}^{\infty}\frac{R_2^b}{R_i^b} = \mathbb{E}\left[\sum_{i=2}^{\infty}\frac{R_2^b}{R_i^b}\right]\right) \\
&\overset{(d)}{=} \left[\frac{(1-\alpha)(b-2)}{\alpha(2^{\frac{\delta}{\tau w}}-1)(b+2)}\right]^{\frac{2}{b}}.
\end{aligned}
\tag{B.5}
$$

The expression in (a) cannot be easily evaluated. We rearrange the expression in terms of R_1 and R_2. As the path-loss exponent increases, $\sum_{i=2}^{\infty}(R_2/R_i)^b \to 1$, meaning that the CDF in (b) approaches the CDF of the ratio of R_1 to R_2. In the absence of a straightforward method to access (b), we approximate it to the conditional CDF in (c), where $\mathbb{E}\left[\sum_{i=2}^{\infty}(R_2/R_i)^b\right] = \frac{(b+2)}{(b-2)}$ [68]. Step (d) follows that $\mathrm{F}_{R_1/R_2}(x) = x^2$ for a PPP over $\mathbb{R}^2$ [22]. (A similar approximation for the spatial reliability follows from [22, Corollary 5], where $[(b-2)/(b+2)]^{2/b}$ is replaced with $\mathrm{sinc}(2/b)$).

The expression above can be rearranged with w on the left-hand side to obtain *Lemma* 5.5, i.e.,

$$
w(\alpha,\eta) \approx \frac{\delta}{\tau \log_2\left(1+\left(\frac{1-\alpha}{\alpha\eta^{b/2}}\right)\times\left(\frac{b-2}{b+2}\right)\right)}. \tag{B.6}
$$

Appendix C

Proof of Lemma 5.7

In the multi-operator network sharing model, the SIR of operator i can be expressed as:

$$\mathrm{SIR}_i = \frac{H_{i,s} R_{i,s}^{-b}}{\sum_{j \in \mathcal{I}_i} H_{i,j} R_{i,j}^{-b}}, \tag{C.1}$$

where $H_{i,j}$ and $R_{i,j}$ are the fading and distances corresponding to (a) the serving base station (i.e., $j = s$) or (b) interfering base stations (i.e., $j \in \mathcal{I}_i$, where $\mathcal{I}_i$ is the set of interferers in the network of operator i).

In multi-operator network sharing with single-connectivity, mobiles associate with the base station that provides the highest ratio of the average received power to the average interference in the multi-operator network, which corresponds to the nearest base station of operator

$$\begin{aligned} i &= \operatorname*{argmax}_{l \in 1..k} \left\{ \frac{\mathbb{E}\left[H_{l,j} r_{l,1}^{-b} \middle| \phi\right]}{\mathbb{E}\left[\sum_{j=2}^{\infty} H_{l,j} r_{l,j}^{-b} \middle| \phi\right]} \right\} \\ &= \operatorname*{argmax}_{l \in 1..k} \left\{ \frac{r_{l,1}^{-b}}{\sum_{j=2}^{\infty} r_{l,j}^{-b}} \right\} = \operatorname*{argmin}_{l \in 1..k} \left\{ \sum_{j=2}^{\infty} \frac{r_{l,1}^{b}}{r_{l,j}^{b}} \right\} \end{aligned} \tag{C.2}$$

for a given set of point patterns $\phi = \{\phi_1, .., \phi_k\}$, where ϕ_i describes the locations of the base stations of operator i, and $H_{i,j} \sim \mathrm{Exp}(1)$ (Rayleigh fading).

This way, using the definitions in Equation (5.1) and Equation (5.3) the link reliability can

be expressed as:

$$
\begin{aligned}
\alpha &= \Pr\left(t_\delta(w, \mathrm{SIR}_i) \leq \tau | \phi, i\right) \\
&\overset{(a)}{=} \Pr\left(\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}} \geq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi, i\right) \\
&\overset{(b)}{=} \Pr\left(\frac{H_{i,1}}{H_{i,2}\left(\frac{r_1^b}{r_2^b} + \sum_{j=3}^{\infty} \frac{H_{i,j}}{H_{i,2}} \times \frac{r_{i,1}^b}{r_{i,j}^b}\right)} \geq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi, i\right) \\
&\overset{(c)}{\approx} \Pr\left(\frac{H_{i,1}}{H_{i,2}} \geq (2^{\frac{\delta}{w\tau}} - 1) \sum_{j=2}^{\infty} \frac{r_{i,1}^b}{r_{i,j}^b} \middle| \phi, i, \frac{H_{i,j}}{H_{i,2}} = \mathbb{E}\left[\frac{H_{i,j}}{H_{i,2}}\right]\right) \\
&\overset{(d)}{=} \Pr\left(\frac{H_1}{H_2} \geq (2^{\frac{\delta}{w\tau}} - 1) \min_{l \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{r_{l,1}^b}{r_{l,j}^b}\right\} \middle| \phi\right) \\
&\overset{(e)}{=} \frac{1}{1 + (2^{\frac{\delta}{w\tau}} - 1) \min_{i \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{r_{i,1}^b}{r_{i,j}^b}\right\}}.
\end{aligned}
\tag{C.3}
$$

The expression in (a) cannot be easily evaluated. Notice that the complementary CDF in (b) approaches the complementary CDF of $H_{i,1}$ to $H_{i,2}$ as $(r_{i,1}/r_{i,j})^b \to 0$, which is generally the case for $j \geq 3$ and $b > 2$. This way, we approximate (b) to the complementary CDF of $\frac{H_{i,1}}{H_{i,2}}$ by conditioning the summation in (b) to its expectation, as in (c). This expectation is undefined for i.i.d $H_{i,j} \sim \mathrm{Exp}(1)$ (Rayleigh fading); however, it can be bound by Jensen's inequality and approximated to its lower bound, which follows $\mathbb{E}\left[\frac{H_{i,j}}{H_{i,2}}\right] \geq 1$. Notice that $H_{i,1}$ and $H_{i,2}$ are independent of i because $H_{i,1}$ and $H_{i,2}$ are i.i.d for all $i \in \{1, 2, ..., k\}$. Furthermore, $\sum_{j=2}^{\infty} \frac{r_{i,1}^b}{r_{i,j}^b}$ where $i = \underset{l \in 1..k}{\operatorname{argmin}} \left\{\sum_{j=2}^{\infty} \frac{r_{l,1}^b}{r_{l,j}^b}\right\}$ is equivalent to $\min_{l \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{r_{l,1}^b}{r_{l,j}^b}\right\}$. This way, we can rewrite (c) as in (d). Step (e) follows from $\overline{\mathrm{F}}_{H_1/H_2}(x) = \frac{1}{1+x}$ [18].

Following Equation (5.2) and using the expression obtained above, we can express the spatial

reliability as:

$$
\begin{aligned}
\eta &= \Pr\left(\Pr\left(t_\delta(w, \mathrm{SIR}) \leq \tau | \Phi\right) \geq \alpha\right) \\
&\overset{(a)}{=} \Pr\left(\frac{1}{1 + (2^{\frac{\delta}{w\tau}} - 1) \min\limits_{i \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b}\right\}} \geq \alpha\right) \\
&= \Pr\left(\min_{i \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b}\right\} \leq \frac{1-\alpha}{\alpha(2^{\frac{\delta}{w\tau}} - 1)}\right) \\
&= 1 - \Pr\left(\min_{i \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b}\right\} \geq \frac{1-\alpha}{\alpha(2^{\frac{\delta}{w\tau}} - 1)}\right) \\
&= 1 - \Pr\left(\bigcap_{i \in 1..k} \left\{\sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b} \geq \frac{1-\alpha}{\alpha(2^{\frac{\delta}{w\tau}} - 1)}\right\}\right) \\
&\overset{(b)}{=} 1 - \prod_{i \in 1..k} \Pr\left(\sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b} \geq \frac{1-\alpha}{\alpha(2^{\frac{\delta}{w\tau}} - 1)}\right) \\
&\overset{(c)}{\approx} 1 - \left[1 - \mathrm{sinc}\left(\frac{2}{b}\right) \times \left(\frac{1-\alpha}{\alpha(2^{\frac{\delta}{w\tau}} - 1)}\right)^{\frac{2}{b}}\right]^k
\end{aligned}
\tag{C.4}
$$

for a set of point processes $\Phi = \{\Phi_1, .., \Phi_k\}$. The expression in (a) stems from (e) in Equation (C.3) when it is conditioned to the set of point processes Φ, as opposed to point patterns ϕ, meaning that $R_{i,1}$ and $R_{i,2}$ are now random variables. Step (b) follows that Φ_m and Φ_n are independent for all $m \neq n$, so are $\frac{R_{m,1}}{R_{m,j}}$ and $\frac{R_{n,1}}{R_{n,j}}$. Step (c) follows that $\mathrm{F}_{\sum_{i=2}^{\infty} R_1^b / R_i^b}(x) \approx x^{\frac{2}{b}} \mathrm{sinc}(2/b)$ for a PPP $\in \mathbb{R}^2$ [22].

The expression above can be rearranged with the bandwidth w on the left-hand side. As in *Lemma* 5.7, the total bandwidth required to support the reliability targets α and η in multi-operator network sharing with single-connectivity can be expressed as:

$$
w(\alpha, \eta) \approx \frac{k\delta}{\tau \log_2\left(1 + \frac{1-\alpha}{\alpha} \times \left(\frac{\mathrm{sinc}(2/b)}{1-(1-\eta)^{\frac{1}{k}}}\right)^{\frac{b}{2}}\right)}. \tag{C.5}
$$

Appendix D

Proof of Lemma 5.9

In multi-operator network sharing with multi-connectivity, mobiles connect to a base station of each operator in the multi-operator network, i.e., k connections in a k-operator network. For each operator, mobiles associate with the base station that provides the highest ratio of the average received power to the interference, which corresponds to the nearest base station in networks with a frequency reuse factor of 1. Signals are combined according to selection combining, i.e.,

$$\mathrm{SIR} = \max_{i \in 1..k}\{\mathrm{SIR}_i\} = \max_{i \in 1..k}\left\{\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}}\right\}. \tag{D.1}$$

This way, using the definitions in using the definitions in Equation (5.1) and Equation (5.3),

the link reliability can be expressed as:

$$
\begin{aligned}
\alpha &= \Pr\left(t_\delta(w, \mathrm{SIR}) \leq \tau | \phi\right) \\
&= \Pr\left(\frac{\delta}{w \log_2\left(1 + \max\limits_{i \in 1..k}\{\mathrm{SIR}_i\}\right)} \leq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi\right) \\
&= \Pr\left(\max_{i \in 1..k}\left\{\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}}\right\} \geq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi\right) \\
&= 1 - \Pr\left(\max_{i \in 1..k}\left\{\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}}\right\} \leq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi\right) \\
&= 1 - \Pr\left(\bigcap_{i \in 1..k}\left\{\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}} \leq 2^{\frac{\delta}{w\tau}} - 1\right\} \middle| \phi\right) \\
&\overset{(a)}{=} 1 - \prod_{i \in 1..k} \Pr\left(\frac{H_{i,1} r_{i,1}^{-b}}{\sum_{j=2}^{\infty} H_{i,j} r_{i,j}^{-b}} \leq 2^{\frac{\delta}{w\tau}} - 1 \middle| \phi\right) \\
&\overset{(b)}{\approx} 1 - \prod_{i \in 1..k} \frac{(2^{\frac{\delta}{w\tau}} - 1) \sum_{j=2}^{\infty} (\frac{r_{i,1}}{r_{i,j}})^b}{1 + (2^{\frac{\delta}{w\tau}} - 1) \sum_{j=2}^{\infty} (\frac{r_{i,1}}{r_{i,j}})^b},
\end{aligned}
\tag{D.2}
$$

where $\phi = \{\phi_1, .., \phi_k\}$ is a set of point patterns describing the locations of the base stations in the multi-operator network. Step (a) follows that $H_{i,j} \sim \mathrm{Exp}(1)$ is i.i.d. for all i and j. As in Equation (C.3), we approximate the CDF in (a) to the conditional CDF when $\frac{H_{i,j}}{H_{i,2}} = \mathbb{E}\left[\frac{H_{i,j}}{H_{i,2}}\right]$ to obtain (b).

Following Equation (5.2) and using the expression obtained above, we can express the spatial reliability as:

$$
\begin{aligned}
\eta &= \Pr\left(\Pr\left(t_\delta(w, \mathrm{SIR}) \leq \tau | \Phi\right) \geq \alpha\right) \\
&\overset{(a)}{=} \Pr\left(\prod_{i \in 1..k} \frac{(2^{\frac{\delta}{w\tau}} - 1) V_i}{1 + (2^{\frac{\delta}{w\tau}} - 1) V_i} \leq 1 - \alpha\right),
\end{aligned}
\tag{D.3}
$$

where $\Phi = \{\Phi_1, .., \Phi_k\}$ is a set of point processes describing the locations of the base stations

in the multi-operator network and $V_i = \sum_{j=2}^{\infty} \frac{R_{i,1}^b}{R_{i,j}^b}$. The expression in (a) cannot be easily evaluated. However, being $V_i \in [0, \infty)$, we can bound the product within the CDF as follows:

$$\begin{aligned}\left(\min_{i\in 1..k}\left\{\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i}\right\}\right)^k &\leq \prod_{i\in 1..k}\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i} \\ &\leq \left(\max_{i\in 1..k}\left\{\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i}\right\}\right)^k.\end{aligned} \tag{D.4}$$

This way, we can evaluate the lower and upper bounds of η as follows:

$$\begin{aligned}\eta &\geq \Pr\left(\left(\max_{i\in 1..k}\left\{\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i}\right\}\right)^k \leq 1-\alpha \,\middle|\, \Phi\right) \\ &\overset{(a)}{=} \prod_{i\in 1..k}\Pr\left(\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i} \leq (1-\alpha)^{\frac{1}{k}} \,\middle|\, \Phi\right) \\ &\overset{(b)}{\approx} \left[\operatorname{sinc}\left(\frac{2}{b}\right)\left(\frac{(1-\alpha)^{1/k}}{(2^{\frac{\delta}{w\tau}}-1)\left(1-(1-\alpha)^{1/k}\right)}\right)^{\frac{2}{b}}\right]^k,\end{aligned} \tag{D.5}$$

and

$$\begin{aligned}\eta &\leq \Pr\left(\left(\min_{i\in 1..k}\left\{\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i}\right\}\right)^k \leq 1-\alpha \,\middle|\, \Phi\right) \\ &\overset{(a)}{=} 1-\prod_{i\in 1..k}\Pr\left(\frac{(2^{\frac{\delta}{w\tau}}-1)V_i}{1+(2^{\frac{\delta}{w\tau}}-1)V_i} \geq (1-\alpha)^{\frac{1}{k}} \,\middle|\, \Phi\right) \\ &\overset{(b)}{\approx} 1-\left[1-\operatorname{sinc}\left(\frac{2}{b}\right)\left(\frac{(1-\alpha)^{1/k}}{(2^{\frac{\delta}{w\tau}}-1)\left(1-(1-\alpha)^{\frac{1}{k}}\right)}\right)^{\frac{2}{b}}\right]^k,\end{aligned} \tag{D.6}$$

where steps (a) and (b) follow from independence among Φ_i (and hence V_i) for all $i \in \{1, .., k\}$ and $\mathrm{F}_V(v) \approx v^{\frac{2}{b}}\operatorname{sinc}(2/b)$ for a PPP $\in \mathbb{R}^2$ [22].

Equations D.5 and D.6 can be rearranged with the bandwidth w on the left-hand side, providing us with the bounds for the required bandwidth at each operator. The required bandwidth can be obtained from the geometric mean of these bounds. Thus, the total bandwidth required to support the reliability targets α and η in multi-operator connectivity sharing in a multi-operator network with k operators can be expressed as:

$$w(\alpha,\eta) \approx \left[\frac{k\delta}{\tau \log_2\left(1 + \frac{(1-\alpha)^{1/k}}{1-(1-\alpha)^{1/k}} \times \left(\frac{\mathrm{sinc}(2/b)}{\eta^{1/k}}\right)^{\frac{b}{2}}\right)} \times \frac{k\delta}{\tau \log_2\left(1 + \frac{(1-\alpha)^{1/k}}{1-(1-\alpha)^{1/k}} \times \left(\frac{\mathrm{sinc}(2/b)}{1-(1-\eta)^{1/k}}\right)^{\frac{b}{2}}\right)} \right]^{\frac{1}{2}}, \tag{D.7}$$

as in *Lemma* 5.9.