Intérêt de la communication direct entre équipements mobiles dans les réseaux radio sans fil.

Thomas Varela Santana

université
PARIS-SACLAY

CentraleSupélec

NNT: 2018SACLC083

On the use of Device-to-Device in Wireless Networks

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n° 580 Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat: Réseaux, Information et Communications

Thèse présentée et soutenue à Paris 13^{eme}, le 9 novembre 2018, par

Thomas Varela Santana

Composition du jury :

Tijani Chahed Professeur, Telecom SudParis — RST	Président
Laura Cottatellucci Professeure, Friedrich-Alexander-Universität — EII	Rapporteur
Rachid El-Azouzi Professeur, Université d'Avignon — LIA	Rapporteur
Emilio Calvanese Strinati Directeur Recherche et Innovation, CEA — LETI	Examinateur
Salah Eddine El Ayoubi Professeur Adjoint (HDR), CentraleSupélec — L2S	Directeur de Thèse
Richard Combes Professeur Adjoint, CentraleSupélec — L2S	Co-encadrant de Thèse
Mari Kobayashi Professeure, CentraleSupélec — L2S	Invitée

Thèse de doctorat

... Mamá, Papá, abuela, esta tesis os la dedico a vosotros ...

Abstract—

This thesis studies Device-to-Device (D2D) communication in realistic and challenging scenarios for future wireless systems. In particular, the thesis focuses on *how may D2D communication help other technologies to enhance their performance*. The first wireless scenario is the one of multicasting, used for example in video streaming or common alert message transmission for police, firefighters or ambulances. The second wireless scenario is the critical one of Ultra-Reliable and Low Latency communication (URLLC) expected to be used to avoid cars crashes in the upcoming Vehicular-to-Everything (V2X) context, and also when connecting machines together in environments like connected hospitals, airports, factories (industry 4.0), and last but not least in e-health context in order to enhance medical tele-surgery. The last wireless scenario is the one of User Equipment (UE) group localization in the context of massive Internet of Thing (IoT), where devices are interacting with each other and are mostly confined in local groups, needing local services.

In the multicast channel scenario, where a transmitter wishes to convey a common message to many receivers, it is known that the multicast rate decrease as the number of UEs increases. This vanishing behavior changes drastically when enabling the receivers to cooperate with each other via D2D. Indeed, the multicast rate increases with high probability when the number of receivers increases. This chapter also analyzes the outage rate of the proposed scheme in the same setting. Extensions regarding firstly resource utilization and secondly considering the use of Hybrid Automatic Repeat Request (HARQ) are also analyzed.

Next chapter addresses one of the major challenges for future networks, named URLLC. Specifically, the chapter studies the problem of HARQ with delayed feedback, where the transmitter is informed after some delay on whether or not his transmission was successful. The goal is to minimize the expected number of retransmissions subject to a reliability constraint within a delay budget. This problem is studied at two levels: (i) a single transmitter faced with a stochastic independent and identically distributed (i.i.d.) noisy environment and (ii) a group of transmitters whom shares a collision channel.

Then the chapter that follows provides a cooperative UE mapping method that is highly accurate. Four different channel models are studied in this chapter: Line of Sight (LOS) and Non-Line of Sight (NLOS) for indoor and outdoor environments. The results show significant improvement compared to already existing methods. Identifying the dense local areas in real time and informing the network allows the Base Station (BS) to increase the capacity through highly directive beams, and therefore, avoids the deployment cost of new infrastructure.

Keywords— D2D, 5G, multicasting, URLLC, HARQ, Positioning

Acknowledgment—

Firstly, I would like to express my gratitude to Prof. Mari Kobayashi and Dr. Sofia Martinez Lopez, who supervised me in the first part of my PhD. The doors to their offices were always open. Afterward, they had to give the supervision's responsibility to Prof. Richard Combes and Prof. Salah-Eddine El Ayoubi from whom I have learned a lot. Many thanks to them.

Besides, I would like to thank the rest of my thesis committee: Prof. Tijani Chahed, Prof. Laura Cottatellucci, Prof. Rachid El-Azouzi and Dr. Emilio Calvanese Strinati, who have spent their time and effort reviewing this manuscript.

I would also like to thank all the people in the Orange Labs RIDE team, for the very interesting discussions. It was a pleasure to come every morning at work. In particular, thank you Pierre for providing such a great environment and for caring so much in your colleagues. Thank you Ana Maria and Berna for supervising me during my internship and keeping our good relation during the PhD. Thank you also Eric for giving me the opportunity to present my work in internal seminar where I had a lot of interesting discussions with Orange members' company. Siham, it has been a pleasure to supervise your six month internship, I am proud of your accomplishment. I would also like to give many thanks to my PhD team mates Rita Ibrahim and Stefan Cerović, we start and end this adventure together. Of course I cannot forget Mohamad Maaz who shares all these years by our side. I would not imagine my PhD without these three persons and I would recommend anyone to work with them. Neighboring Orange Labs teams were also full of incredible PhD students: muchas gracias Felipe, shokran ktir Tony e in particolare grazie mille Julien who also shared L2S laboratory with me at CentraleSupélec, sei fantastico.

I would like to thank Jose Fonseca, Huu-Hung Vuong, Catherine Magnet and Anne Batalie for making the technical and administrative procedures easier.

Évidemment je ne pouvais pas finir cette section sans remercier mes amis, d'enfance, ceux rencontrés durant mes études, qu'ils soient en région parisienne ou bien plus éloignés, nos rencontres ont sûrement diminuées durant ces années, mais ma reconnaissance à votre égard n'a fait qu'augmenter. Je remercie également ma famille proche qui m'a soutenue tout du long, Rémy tu as du supporter deux proches en thèse, quel héros tu es. Queria dar gracias también a esa persona, quien añadió felicidad y me hizo empezar mis días con alegría. A mi familia de Santa Pola y Vilapouca, este año no pude unirme a vosotros, que conste que no volverá a pasar. Por último, como no basta con tener diplomas para tener buena educación [1], gracias a mis padres por haberme ofrecido la mejor posible.

Notations— Here we define the mathematical notations used all along the thesis

x	scalar
$\mathbf{x}$	vector
$\mathbf{X}$	matrix
X	random variable
(X_n)	sequence of random variables
$\triangleq$	is defined as
$\sim$	is distributed as
$\gg$	much greater than
$\mathbb{R}$	the set of real numbers
$\mathbb{R}_+$	the set of positive real numbers
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$[a, b]$	the close interval $\{x \in \mathbb{R} : a \leq x \leq b\}$
(a, b)	the open interval $\{x \in \mathbb{R} : a < x < b\}$
$\|.\|$	cardinality of a set or absolute value of a scalar
$\lceil . \rceil$	Ceiling function
$\lfloor . \rfloor$	Floor function
$a \mod b$	a modulo b, $a - (\lfloor a/b \rfloor b)$
$e^{(.)}, \exp\{.\}$	exponential function
$\ln(.)$	natural logarithm function
$\log_a(.)$	logarithm to base a, $\ln(.)/\ln(a)$
$\mathbb{1}\{.\}$	indicator function
$\mathcal{O}(.)$	Landau notation
$C_n^k, \binom{n}{k}$	binomial coefficient $\frac{n!}{k!(n-k)!}$ for $n, k \in \mathbb{N}$ and $k \leq n$
$\mathbb{P}(.)$	probability
$\mathbb{E}(.)$	expectation
var(.)	variance
$f(.)$	value of function f at .
$\lim_{a \to b} f(a) = \bar{f}, f(a) \underset{a \to b}{\to} \bar{f}$	$\bar{f}$ is the limit of $f(a)$ as a tends to b
$\limsup_{a \to b} f(a)$	limit superior of $f(a)$ as a tends to b
$X_n \xrightarrow{\mathbb{P}} X$	X_n converges to X in probability
$X_n \xrightarrow{d} X$	X_n converges to X in distribution
$X_n \xrightarrow{L^1} X$	X_n converges to X in mean
$X_n \xrightarrow{L^2} X$	X_n converges to X in mean square
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2

Chapter 1

Introduction

1.1 Motivation

In the context of future wireless networks, the Fifth-Generation (5G) is expected by 2020[1]. While the Fourth-Generation (4G) is mostly about ensuring a constantly growing and changing capacity requirements, 5G network can be imagined as a software infrastructure where the goal will be to deliver a wide variety of services. From a technical point of view, 5G network will have to answer to the three major classes of services: enhanced Mobile BroadBand (eMBB), massive Internet of Thing (IoT) and Ultra-Reliable and Low Latency communication (URLLC), as described in Fig. 1.1. eMBB is the continuity of the multimedia-based services provided by the 4G, where the main requirements concern mostly high data rates and extensive coverage. To meet those requirements, carrier aggregation, new cm/mmWave bands and massive Multiple-Input Multiple-Output (MIMO) among others have been proposed. Regarding the massive IoT service, full coverage with a high density of cheap devices is proposed for services like smart cities. For URLLC, the unprecedented reliability and low latency requirements open the door to services like connected cars, industry 4.0, e-health, smart cities, immersive gaming among others.

All the previous scenarios imply new demands and give us the opportunity to deal with different challenges such as the high number of devices, long range coverage, multimedia mobile video streaming, low latency, high reliability, position accuracy for indoor and outdoor.

[1] https://www.orange.com/en/Human-Inside/Thematic-feature/5G-the-mobile-connectivity-of-the-future.

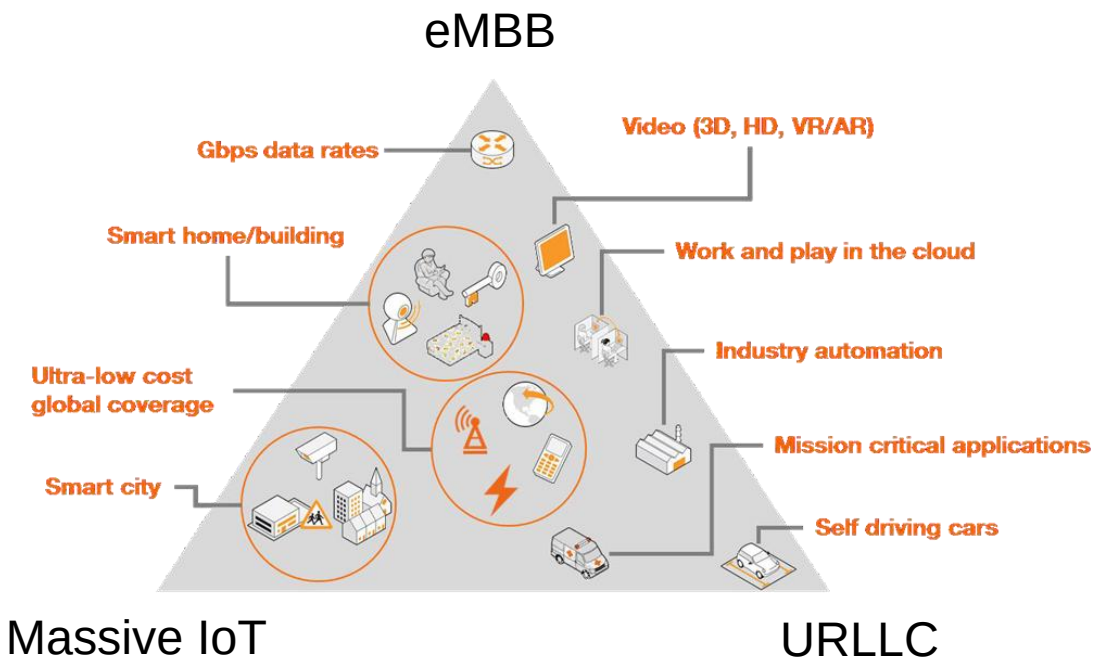


Figure 1.1: 5G scenarios [2].

One of the envisioned technologies in order to meet all the cited challenges and that is part of the 5G cellular network is Device-to-Device (D2D) communication. This technology offers direct communication between the User Equipments (UEs), enabling new 5G business models. D2D connectivity could be, for an operator, a way to offload the traffic from the network. Exploiting direct communication enables Proximity Service (ProSe), which is useful for connected vehicles (URLLC), dense networks (massive IoT), but also for coverage extension (eMBB). It improves spectrum utilization, overall throughput, latency and can be used for location-based applications.

In particular, in this thesis, we explore several wireless network scenarios, represented in Fig. 1.2, where D2D brings a fundamental added value. Description and contributions of these scenarios are summarized in the following sections.

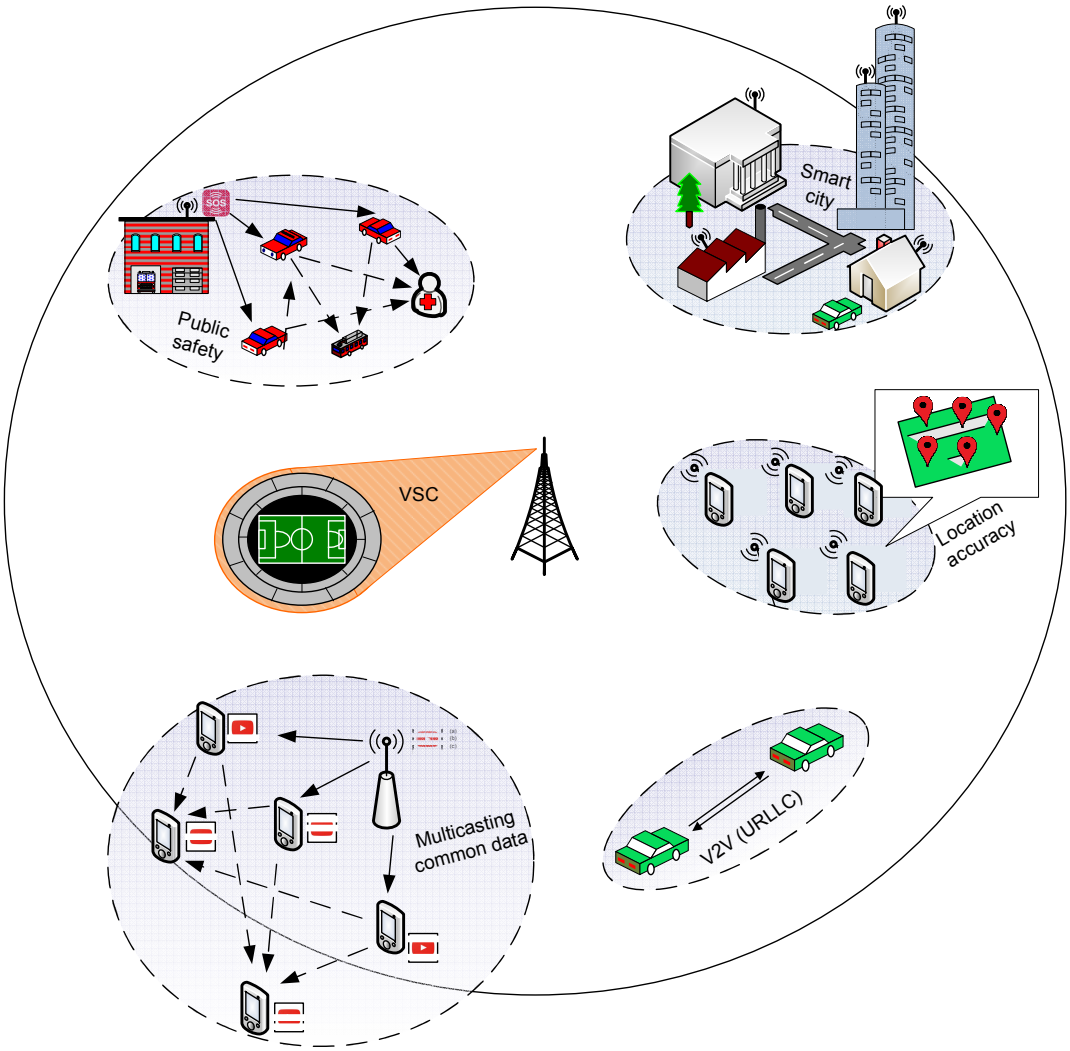


Figure 1.2: Overview of the studied scenarios.

1.2 Thesis Outline and Contribution

This thesis will focus on the three following problems:

1.2.1 Multicast Wireless Problem

A first important wireless use case is multicast communication, where a single transmitter sends a common message to many receivers in the presence of fading. There are two ways to measure the efficiency of a communication protocol:

- The average multicast rate, which is the expected number of successfully decoded bits per channel use achieved by a UE chosen uniformly at random;
- The multicast outage rate, which is the maximum rate at which all the UEs can decode the message with an error probability of ϵ.

Firstly, the average multicast rate is interesting in the context of wireless edge caching. It has been shown that the traffic during the peak hours can be significantly reduced by caching popular contents during off-peak hours, and delivering these popular contents using multicasting (see e.g. [3, 4] and references therein). The problem when studying the average multicast rate is that it remains constant in dense scenario, i.e. as the number of UEs grows.

Secondly, the study of the multicast outage rate is essential for the so-called enhanced Multimedia Broadcast Multicast Service (eMBMS) scenario [5]. These services reduce the network load by multicasting common data to Public Safety (PS) devices [6], where decoding is mandatory for every device. Here the multicast outage rate is limited by the UE in the worst channel condition and hence the multicast outage rate vanishes in the presence of large number of UEs.

In this first contribution, motivated by these two scenarios, we study D2D-aided multicasting without Channel State Information at the Transmitter (CSIT) to overcome the limited rate achieved by the multicast communication. Instead of considering CSIT, we only suppose statistical knowledge of the channel given by the receivers each period of time. This avoids channel measurement overhead and also error measurement in case of fast fading. We consider a general network topology combined with D2D technology, where the transmitter has only statistical channel knowledge. Furthermore, we analyze the multicast rate in a dense network, i.e. in the regime of large number of UEs. In this setup, we address the fundamental question: *can D2D without CSIT increase the achievable multicasting rate?*

Our contributions for the multicast wireless problem are the following:

1) We propose a simple two-phase scheme that achieves an increasing average multicast rate while requiring only statistical channel knowledge at the transmitter and local channel knowledge at each receiver. The enhanced two-phase scheme with the optimal resource split achieves the average multicast rate of order $\mathcal{O}(\ln \ln n)$ with high probability, where n is the number of UEs.

2) We derive tractable asymptotic expressions for both the average and outage multicast rate in the asymptotic regime of a large number of UEs. It is worth noticing that the derivations of such deterministic equivalent expressions are novel and involved. They are based on non-trivial concentration arguments.

3) We further provide approximate expressions for both the average and outage multicast rate. Our numerical examples show that these expressions are very accurate even for a small and finite number of UEs (say $n \geq 50$).

1.2.2 URLLC Wireless Problem

A second important wireless use case is represented by URLLC. URLLC requires ultra-high reliability and low latency transmission. This is even more challenging because reliability and latency constraints are usually not considered in the same system. Using D2D communication under URLLC constraints enables us to introduce services like connected cars (that use an extension of D2D named Vehicular-to-Vehicular (V2V) or Vehicular-to-Everything (V2X)), but also for services like e-health or industry 4.0 when connecting machines together, i.e. Machine-to-Machine (M2M) communication.

In URLLC systems, packet transmission time is constraint by low latency, implying packet deadline and short time slots for transmission. In this way, 3rd Generation Partnership Project (3GPP) starts standardizing short time slots (e.g. 1 ms in Long Term Evolution (LTE), 500 μs and 125 μs in New Radio (NR)) [7, 8]. This is due higher carrier sub-spacing, possible thanks to wider bandwidth (20 MHz in 700 MHz carrier frequency for LTE, 100 MHz in 3.5 GHz carrier frequency and 400 MHz in 28 GHz for NR, respectively with the time slots [7]).

Now, in order to face the high reliability constraint of URLLC, a retransmission procedure like Hybrid Automatic Repeat Request (HARQ) is mandatory (actually, in order to reach a high reliability in the presence of noise, it is impossible to avoid retransmission). In a D2D communication using HARQ, the UE transmitter is waiting for the Acknowledgment (ACK) feedback from the UE receiver before retransmitting. This delay is not affordable and negligible, even more with such shorter time slots as mentioned previously, in many use cases like the cited ones. Reliability is achieved thanks to re-

transmissions, hence waiting for the feedback may fail to transmit the packet enough times before the deadline constraint, canceling the URLLC requirement.

In that sense, the approach consisting in combining "wait the Non Acknowledgment (NACK) before retransmission" and "blind retransmission" needs to be implemented and optimal policy has to be derived. In other words, the approach consisting in optimally retransmitting the same packet without necessarily waiting for the NACK is studied under reliability and latency constraint.

Our contributions to the design of URLLC are the following:

1) We analyze the optimal HARQ schemes for URLLC services when the feedback arrives with a delay that is larger than the transmission interval (as it is usually the case). We derive the optimal policies and show that they can be computed efficiently.

2) We propose a novel policy that has the advantage of being simple; of relying only on the knowledge of the service requirements and on the channel feedback delay; of not depending on the link error rate; and of being optimal in many practical cases that we identify.

3) We extend our results to the system level where there are multiple transmitters competing for the channel access and requiring URLLC service. We derive retransmission policies in order to meet the URLLC targets with the smallest number of retransmissions.

While the solutions we propose apply to many URLLC use cases, including UE to BS ones, they are particularly relevant for URLLC involving D2D. Indeed, the distributed nature of URLLC D2D makes centralized scheduling difficult and calls for contention-based approaches. NACKs in this case may be implicit, i.e. the absence of an ACK may be interpreted as a NACK as the receiver may not be aware of the existence of the packet.

1.2.3 Localization Problem for Wireless Systems

Finally, with the increasing number of IoT devices, services like smart city appear. Usually in such a system the devices are interacting with each other's, and indoor/outdoor positioning needs to be very accurate. Being able to estimate location with high accuracy is also important in services like self-driving (avoid car crashes), e-health (surgery or person assistance at home), smart home (domotic), social networking (finding friends located close to one) and industry 4.0 (packet warehouse). Notice that the equipments used in smart cities or industry 4.0, also called smart devices, need to be of low cost, low energy consumption and the transmission between them needs to be robust to cellular

interference and to fading. Because it respects all the previous requirements and because it has been standardized by the Institute of Electrical and Electronics Engineers (IEEE) 802.15 working group, which specifies Wireless Personal Area Network (WPAN) standards, we propose to consider D2D communication with Ultra-WideBand (UWB) signals between the devices. UWB emits short low energy pulses spread in a very large bandwidth, providing very high accuracy due to its wide frequency band.

As the devices in smart city or industry 4.0 tend to be mostly confined in local groups, it is of great interest for the local network to know their location and cluster them with highest possible accuracy. Cooperative localization of UEs is hence a key feature for 5G.

Our contributions in this third wireless network are the following:

1) In this third work we propose a dynamic cooperative equipment mapping using UWB signals, where map accuracy is studied in LOS and NLOS environments. This method provides an accuracy of cm order.

2) A comparison of our technique with Global Positioning System (GPS) is also presented.

3) This method is applied to detect hotspots. In addition, in order to avoid small cells deployment, a large number of antennas at the BS can be exploited, creating highly directive beams to cover a given area. This concept is called Virtual Small Cell (VSC) [9]. This coverage scenario is useful in 5G services like ultra-low cost global coverage, smart cities or industry 4.0.

1.3 Organization

The remainder of this thesis is organized as follows: in Chapter 2 we present prior work relevant to this thesis work, including wireless networks, D2D communications, multicast channels, URLLC and device localization. In Chapter 3 we study D2D aided multicasting, where two metrics of interest are derived (average and outage rate). We also extend the analysis by taking into account resource utilization. A second extension with HARQ scheme at the UE level is also studied. In Chapter 4 we propose a new URLLC transmission scheme, where we derive the optimal retransmission policy subject URLLC constraints for single D2D pair and also for multi UE transmission. In Chapter 5 we study massive IoT network where we propose a highly accurate mapping method. This chapter also introduces the VSC concept with a direct application of the mentioned localization method. Finally, Chapter 6 concludes the thesis and presents the possible directions for future work.

Chapter 2

State of The Art

2.1 Device-to-Device Fundamentals

D2D communication is defined as a direct communication between two UEs without routing through the BS. A key motivation for D2D connectivity is the potential for an operator to offload traffic from the network, extending the cellular coverage and facilitating new types of wireless ProSes like hyper-local advertisement [10]. D2D can also reduce costs for operators as it does not need another cellular infrastructure than the already existing one. On top of being a cost reduced technology, D2D communication will still be available when cellular network fails due to, for example, natural disaster like earthquake, which is interesting for Public Safety (PS) applications.

From [11] the first 3GPP release introducing D2D is Release 12, starting D2D standardization in 2012. The main motivation for adding D2D to 3GPP was the use case of ProSe. In ProSe, the needs are to exchange data and voice among nearby UEs (which introduce D2D communication) and the discovery of UEs or services (introducing D2D discovery). In Release 12, D2D discovery is studied only for the in-coverage scenario, i.e. for UEs under BS coverage. The D2D communication in Release 12 is limited to broadcast communication in-coverage and out-of-coverage scenarios. A second important use case that push D2D standardization is the one of PS. PS proposes to use ProSe within the broadcast communication for applications such as police, firefighters and ambulances. D2D enhancements are proposed in Release 13, which started in 2014, where one-to-one communication has been introduced in the standards, and where discovery has been extended to partial and out-of-coverage scenario. A third use case of ProSe UE-to-Network Relay has been proposed by 3GPP organization, where a UE act as a relay between a remote UE and the BS [12]. Since Release 14 [13], D2D has been replaced by V2X (including V2V, Vehicular-to-Infrastructure (V2I), Vehicular-to-Network (V2N) and Vehicular-to-Pedestrian (V2P)), which aim to develop a vehicular communication system able to guarantee safety applications as well as comfort driving services with

a high degree of quality and reliability (more details are provided in Section 2.3.2). Regarding Release 15, the main use cases considered among others are platooning (operating a group of vehicles in a closely linked manner so that the vehicles move like a train with virtual strings attached between vehicles), extended sensors (enables the exchange of live video images among vehicles, road site units, pedestrian devices and V2X application servers), advanced driving (enables semi-automated or full-automated driving) or remote driving (vehicle is controlled remotely by either a human operator or cloud computing) [14]. The main use cases for D2D/V2X in the different releases of 3GPP are summarized in Table 2.1.

Table 2.1: Main use cases for D2D in the different Releases of 3GPP

Release 12	Release 13	Release 14	Release 15
PS	PS enhancement	V2V	Platooning
ProSe	ProSe enhancement	V2I	extended sensors
	UE-to-Network Relay	V2N	advanced driving
		V2P	Remote driving

Going back to Release 12 and 13 that deal with D2D, all the studied ProSe use cases depend on D2D communication and discovery. 3GPP studies mainly focused on the technical details, including radio resource allocation and management, like spectrum utilization and power control, or D2D procedures like synchronization, discovery signal design, channel quality estimation, relaying etc.

2.1.1 Device-to-Device resource allocation and management

D2D communication can be spectrally divided into two parts (see Fig. 2.1), namely inband D2D communication, where both D2D and BS use the cellular band, and outband D2D communication, where D2D uses a dedicated band [15, 16]. Regarding dedicated radio communication, IEEE continue studying Wi-Fi Direct within the IEEE 802.11 Study Group and also Ultra-WideBand (UWB) within the IEEE 802.15 Working Group which specifies WPAN standards for mobile social networks [17].

2.1.1.1 Inband Device-to-Device communication

The communication under this category happens in licensed spectrum (i.e., cellular spectrum) and is used for both sidelink and cellular link, where sidelink is the name of D2D link given by the 3GPP. The motivation for choosing inband communication is usually

the BS control over cellular spectrum (i.e., licensed band) [18, 19, 20]. Still the interference in the licensed spectrum is hard to manage and imposes constraints for quality of service provisioning [21].

Inband D2D is further divided into underlay and overlay categories.

1) In underlay D2D communication, cellular and D2D communication share the same radio resources. Underlay D2D increases the spectral efficiency of cellular spectrum [22], whereas the interference management among D2D and cellular transmission in underlay is very challenging [16];

2) In contrast, sidelinks in overlay communication are given dedicated cellular resources. In overlaid mode, D2D communications operate in licensed spectrum, but remain completely transparent to the cellular UEs [16]. The disadvantage of overlay mode is that cellular resources might be wasted;

D2D inband communication uses cellular spectrum and promises different types of gain: (a) the UEs proximity will provide extremely high bit rates, low delays, and low energy consumption [23, 18]; (b) cellular spectrum can be fully controlled by the BS implying that radio resources may be simultaneously used by cellular links as well as sidelinks. Same spectral resource can hence be used more than once within the same cell [23]; (c) D2D technology can be used by any cellular device;

The disadvantages of inband D2D communication are: (a) interference management solutions usually resort to high complexity resource allocation methods [24]; and (b) in the case of underlay, a UE cannot have simultaneous cellular transmissions and D2D communication [25].

One of the reasons why underlay D2D communication is more considered in the literature than overlay, is because allocating dedicated spectrum resources to D2D UEs in overlay is not as efficient as underlay in term of spectral efficiency [26].

2.1.1.2 Outband Device-to-Device communication

The D2D communication under this category exploits dedicated band or unlicensed spectrum. The motivation behind is to eliminate the interference between sidelinks and cellular links, since outband D2D communication does not occur on cellular spectrum [16]. Using dedicated band requires an extra interface and usually adopts other wireless technologies, such as Wi-Fi Direct, ZigBee, Bluetooth, LTE-U (unlicensed) or UWB [27]. In other words, only cellular devices with two wireless interfaces (e.g., LTE and Wi-Fi/UWB/Bluetooth) can use both cellular and dedicated bands, and thus have simultaneous D2D and cellular communication [28, 29].

The literature suggests to divide outband D2D in two type of communication [30, 31, 32, 33].

1) Controlled communication, the control of second interface/technology is under cellular network [25, 26];

2) Autonomous communication, cellular communication is controlled by the network, but leaves all the D2D communication to the UEs (second interface/technology is not under cellular control) [25];

D2D outband offers to eliminate the interference between D2D and cellular communication, and offers also the possibility to have D2D and cellular communication at the same time. Whereas the disadvantages of D2D outband communication are an uncontrolled unlicensed spectrum and it can only be used by the cellular UEs with two radio interfaces [26].

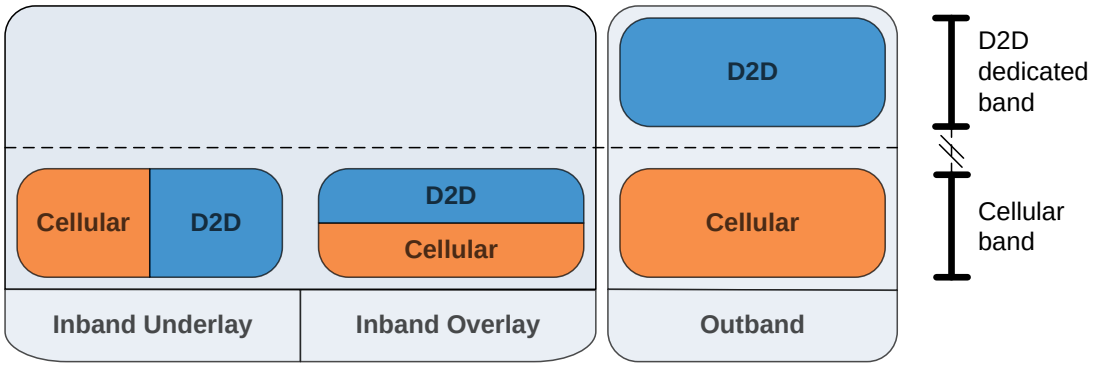


Figure 2.1: Summary of the D2D bands.

2.1.1.3 Devive-to-Device radio resource management

It is known that PS UEs usually have access to dedicated spectrum (unlicensed band); However, commercial D2D UEs have to share the radio resources with cellular UEs in Frequency Division Duplexing (FDD) or Time Division Duplexing (TDD) networks [18]. Hence, 3GPP ask the question: which part of the radio resources should D2D inband transmission utilizes: downlink, uplink or both link resources? When D2D transmission utilizes downlink resources, a transmitting D2D UE may cause high interference to nearby co-channel cellular UEs receiving downlink traffic. Uplink resources are often less utilized compared to the downlink, and sharing the uplink resources with D2D may improve spectrum utilization. In case of inband underlay, the receiving D2D UE will experience interference from nearby co-channel cellular UEs transmitting uplink traffic [24]. Finally, note that reusing uplink resources in FDD requires that the UE is capable of receiving in the uplink.

In case of in-coverage UEs, for inband and outband controlled communication, the network will manage the radio resource. Usually part of the resources are periodically allocated, for D2D discovery (semi-static allocation), or dynamically allocated [34], based on current transmission demands in the case of D2D communication, as D2D traffic may vary significantly. This is called scheduled allocation mode, where D2D transmitter asks for a D2D transmission request to the BS. D2D signals are then transmitted in the resources indicated by the BS. Also, the network can allocate a resource pool dedicated to D2D communication and let the D2D UEs choose the adequate resource allocation by themselves, called autonomous allocation mode, using random access protocols (e.g. Carrier Sense Multiple Access (CSMA) to deal with the eventual collisions). Autonomous resource allocation is also used in out-of-coverage scenario [35] where the resource pool can be pre-configured with for example a distributed resource access protocol [36]. Finally, autonomous allocation can be done in in-coverage and out-of-coverage scenario whereas scheduled allocation can be done only in in-coverage scenario.

2.1.1.4 Device-to-Device power control

D2D communication can work in both fixed power scheme and fixed Signal to Noise Ratio (SNR) target scheme [37]. In the fixed transmission power case, all UEs in D2D mode, use the same transmission power. This scheme is simple, but during the synchronization phase (in the D2D outband autonomous) where the resource randomization procedure is made, it does not work well as there can be significant interference. In the fixed SNR target case, the selection of the SNR target will affect the total transmission power and the final Signal to Interference Noise Ratio (SINR) directly [38, 39, 37]. More transmission power for D2D UEs will improve the final SINR for those D2D UEs. However, there could be a risk of increasing the overall interference level to the UEs of the cellular mode in case of D2D inband underlay communication [40]. In both schemes, the maximum allowed power is set to 23 dBm with an exception for PS devices where transmission is allowed up to 31 dBm in the 700 MHz band [34].

2.1.2 Device-to-Device procedures

2.1.2.1 Device-to-Device synchronization

In order to demodulate the D2D data, the UE receiver has to be synchronized in time and frequency to the UE transmitter. In case of coverage scenario (i.e. D2D transmitter and D2D receiver are under the coverage of the same cell or of two synchronized cells) the synchronization is provided by the BS, which one should inform the D2D UEs what time slot and frame timing as well as frequency synchronization they are going to use [41]. If the UEs are in out-of-coverage scenario, synchronization is made by using some randomized procedure like a UEs sending/scanning for beacons at different times and frequencies

without knowing the physical location of the intended UE until the moment they discover one with high reliability [18]. Here UEs need to be continuously active, which leads to high energy consumption [42]. This is the approach used in Bluetooth, Wi-Fi Direct or UWB [43]. If one UE is in out-of-coverage scenario, or if UEs are in different non-synchronized cells, then network assistance is available, and the in-coverage UEs will transmit the synchronized signal to the others. In case of several synchronization signals received by a UE from differently synchronized UEs, the selection is based on the closest UE to the BS, implying higher reliability [44]. Therefore, the D2D UEs can be assumed to be synchronized to each other prior to sidelink establishment.

2.1.2.2 Device-to-Device discovery and channel quality estimation

The first step in the establishment of a sidelink is the discovery of available devices. Device discovery is made possible by one UE sending a synchronization/reference (beacon) signal and other UE scanning for and capturing such a beacon [45, 46]. Please note that some scenarios like proximity advertisement only need D2D discovery. From a resource management point of view, semi-static allocation may allow the UEs involved in discovery to save energy [47], for example by sleeping let us say for 99% of the time and only wake up to transmit/receive discovery signals in the predefined subframes. By measuring and reporting the quality of the received beacon signals, the device and the BS can estimate the quality of the D2D channel and thereby determine whether a sidelink can be established or not [11]. During the D2D discovery process, one UE tries to discover the UE that can provide him the service he is interested in [48]. In practice, these UEs can swap the roles at different occasions, or simultaneously. D2D discovery may be classified into two categories: restricted discovery and open discovery. In open discovery, no explicit permission is needed from the UE being discovered, whereas restricted discovery, only happens with explicit permission from the UE being discovered [49].

2.1.2.3 Device-to-Device relay

In order to extend the coverage area, D2D relaying has been considered. The remote UE (the one who needs better connection) can connect to the relay UE in several situations: when it is out-of-coverage, when battery saving is necessary (less transmission power is needed for D2D communication), when it is in-coverage, but at the border and wants to benefit from a neighboring cell connection, etc. [50]. The first step of the relay procedure is initiated by either the BS or by an in-coverage UE. Once an in-coverage UE has been selected for relaying, it starts the D2D discovery procedure. The third step is to establish a connection between the relay and the UE discovered (remote UE). Resources are then allocated and a communication between the remote UE and the BS through the relay UE is available [44].

Now that the general concepts of D2D have been presented, an overview of the three main features of this thesis (multicasting, URLLC and localization) with and without D2D is needed.

2.2 Multicast

2.2.1 General state of the art

In a huge and diversified number of scenarios, transmitting common information to many receivers could be useful. In such a multicast scenario, scheduling packets from a number of data streams into groups has been deeply studied in [51]. Still, considering groups of UEs, the authors in [52] propose an approach for a multicast throughput analysis. Some works like [53] are considering CSIT at the BS before multicasting the common message instead of studying a "blind" transmission scenario. The capacity of multicast channel has been extensively studied in the literature [54, 55] as well as scheduling techniques [56]. Since the common message must be decoded by all receivers, the capacity of the fading multicast channel is limited by the UE in worst condition. In the case of the i.i.d. Rayleigh fading channel, it is known that the multicast capacity vanishes inversely to the number n of UEs as n grows [54, 57]. Despite the multicast vanishing rate, this transmission is relevant to two scenarios in wireless crowded networks. The first scenario is the wireless edge caching and applications like local file transfer in commercial networks. Reduction of the traffic during peak hours is possibly done by caching popular contents during off-peak hours, and transmitting using multicasting [3, 4, 58]. A novel UE selection scheme requiring only statistical channel knowledge has been proposed for the i.i.d. Rayleigh fading channel [59] as well as for the asymmetric Rayleigh fading channel [60]. The second scenario is the group communication, standardized by the 3GPP [61], for eMBMS aided by ProSe. These services reduce the network load by multicasting common data to PS devices [6].

2.2.2 Device-to-Device for multicast

Still with the objective of improving the multicast channel, several works start introducing and combining D2D technology with multicasting. Consider the case where multiple UEs are requesting the same contents from the BS. It could be video streaming of most popular programs, for instance, during world cup, when multiple UEs are watching the same football match. These UEs can first form cooperative clusters according to the geometry (see an example of clustering in Chapter 5), to achieve a higher energy and spectrum efficiency. In a first step, BS can transmit the common contents to some UEs in good conditions. In a second step, each UE that successfully decode in the first step will multicast the contents to other UEs within the cluster through sidelinks. Please

note that the BS can stay silent during the second step and hence keep the network energy efficient. This kind of communication can improve the system throughput in case of systems that schedules the best UE with the best channel quality. If there are several channels, the probability that all channels are deteriorated at the same time is quite small. Hence this D2D relaying communication will increase reliability.

In order to support group communication as well as multicast for ProSe, solutions for one to many communications need to be developed. This requires the serving operator's authorization prior to communication. A first work considering D2D multicasting above a cluster in order to perform a coverage analysis is [62]. In this work, a Cluster Head (CH) that already received the message from the BS is retransmitting the content to the UEs in his cluster by sidelinks during T times. The BS that benefits from CSIT will help by retransmitting at most once. This enables to increase the coverage probability, as well as the mean number of covered UEs.

Other improvements like interference can be performed thanks to D2D. This is the main objective of [63, 64] where authors consider several cellular UEs and D2D UE multicasting groups, sharing the same frequency band than the cellular ones (underlay). They consider a cellular SINR threshold and a maximization of the D2D power transmission under which the BS, with CSIT, will propose an adequate resource allocation. This enables to decrease the interference and hence increase the throughput.

In [65], D2D multicasting is used as a clustering scheduling based on the outage probability of the D2D UEs. Depending on this probability, UEs will belong to the cluster or not. In other words, if a UE is in outage with high probability, it will not belong to the cluster. This UE selection will form a cluster of UEs in good conditions.

Also, in [66] it is proposed to improve multicast reliability thanks to D2D and HARQ feedback. Indeed, the BS will transmit the common message at a given rate. The UEs that successfully decode the cellular message will send an ACK, and the ones that unsuccessfully decode the cellular message will send a NACK, to a predefined CH. This avoids feeding back all the ACKs/NACKs to the BS and overloading the uplink. Once the CH collects all the ACKs/NACKs, he will classify all the feedback messages in four cases: All_ACK, All_NACK, $Self\_ACK$ (CH successfully decodes but one UE did not) and $Self\_NACK$ (CH did not successfully decodes but one UE did). Such a feedback scheme can be sent to the BS in 2 bits, reducing the uplink overload. The BS will then multicast next message with adapted rate and smaller error probability. This D2D scheme will decrease the overhead consumption and increase the energy saving for cellular links.

Authors in [67] also use HARQ and cooperative retransmission in order to solve the multicast issue of suffering from the worst UE link. The first step is that the CH (or BS) transmits a common message to cluster members. During the second step, UE members report the eventual NACK also by "blindly" multicasting. In the meantime, all the

UEs that successfully decode turn on their feedback listening mode by supposing HARQ synchronization. Finally, in step 3, all the UEs that heard the NACK retransmit data to the feedback senders. This combination of multiple retransmissions provided by the UE members will increase the success probability of the UEs sending a NACK.

Still in the scheme of one BS multicasting a common message to multi D2D-devices, [68] proposes a relay selection of the UEs that successfully decode in downlink based on the maximization of the resource utilization. This relay selection enables a resource efficient system.

A multicast two-phase scheme, where the first phase is reserved for the downlink communication and the second phase for sidelinks between the UEs that successfully decode and the ones who did not, has been studied in [69]. In this work, the authors consider that the channel statistics of all the UEs are symmetric and, their metric is on the error probability decay by exploiting channel hardening via space-time code.

2.3 Ultra Reliable and Low Latency Communication

2.3.1 General state of the art

A new class of services is emerging for wireless networks, characterized by stringent requirements on latency and reliability. These services are classified under the URLLC class [70]. These services are in general related to industrial IoT applications like in factories 4.0 or in smart hospitals. Typical requirement is that a large proportion of packets (99,999%) need to be received within a latency budget of 1 ms [71]. Packet retransmission is a key enabler for improving the reliability performance [72], but using classical HARQ retransmission procedures introduces additional latency [72] and other retransmission schemes are needed for URLLC.

The design of retransmission schemes has been a hot research topic in the recent years. For instance, authors in [73] investigated joint link adaptation and HARQ to maximize the spectral efficiency under delay and error performance constraints. The authors in [74] derive the optimal number of retransmissions so that the spectral efficiency is maximized while achieving the latency and reliability targets. These works consider that retransmissions always happen after reception of a NACK, which may not be suitable for very low latency budgets. Another set of works adopts a different approach, with blind retransmissions before any ACK/NACK is received, targeting a low latency. In this regard, an approach that consists in sending multiple replicas of the same packet without waiting for the acknowledgments is already adopted as a solution in the 3GPP standard [75]. The work in [76] proposes to send these replicas in consecutive Time Transmission Intervals (TTIs), where the resources used by each replica are randomly selected from the set of available Resource Blocks (RBs) in each TTI. Such schemes

are generally associated with a "grant-free" approach, under which neither issuing a scheduling request nor waiting for a scheduling grant are required [77]. A contention-based access is the appropriate scheme in this case, where the UEs contend to access some shared time and frequency resources which are preallocated for the contention procedures [78]. This results in possible collisions between packets, generating losses that are added to the losses due to imperfections in the radio channel. In this context, the number of packet replicas has to be optimized so that the reliability targets are achieved, and the optimal policy can be derived for a known number of UEs and when all UEs follow the same policy [76].

2.3.2 Device-to-Device for Ultra Reliable and Low Latency Communication

One important use case requiring strict delay communication is the vehicular traffic safety. Vehicular communication is named as V2X, and two complementary transmission modes are defined in the standards [79]. The first mode is the communication between two vehicles through cellular link and the second one is direct communication through sidelink. Here we will be interested in the second mode, where V2V transmission is nothing more than an extension of D2D for vehicles with strict delay requirements transmitting data packets with information about its location, speed, direction, etc. [80]. Low latency and reliability are essential for collision avoidance systems, for example to coordinate braking between vehicles for autonomous safety systems. By using D2D rather than relying on cellular infrastructure, it is possible to firstly reduce the latency due to close proximity. Still the use of D2D is not enough to meet neither the vehicular latency requirement of ms order, nor the transmission reliability of $99,99\%$. A technology that increases the reliability is HARQ due to its possibility of retransmitting until the receiver successfully decodes. In order to meet these requirements, 3GPP standards have decided to combine D2D and HARQ technology for vehicular transmission (similar requirement for e-health like surgery is needed). It is known that latency will be added using HARQ [72], but with the upcoming new generation 5G network, bandwidth is supposed to be larger, hence TTI will be shorter, implying a latency reduction. This is one of the reasons why platooning use case has been introduced in 3GPP Release 15 [14] and enhanced in Release 16 [81]. Platooning is the coordination movement of several vehicles in close proximity. Note that the 5G V2X study item includes licensed frequencies up to 52.6 GHz for sidelinks [82]. In addition, new HARQ retransmission design (as deeply described in the previous Section 2.3.1) can be used.

A second important use case requiring strict delay communication is the industry automation, also called industry 4.0. This use case includes more generally connected factories, warehouses, hospitals, airports, etc. In order to avoid the deployment cost of new physical architecture like small BSs, direct connection between the objects/machines

is needed [83]. Avoidance of Capital Expenditure (CAPEX) by using D2D communication is also justified by the demo of a product for example in temporary and external sites. Such a use case needs to support licensed and unlicensed spectrum (see Section 2.1.1) and require high performances in term of reliability and latency. A typical traffic model for industry 4.0 is a packet size of 100 bits, an average package rate of 0.1 packets/ms (per machine), a machine density of 0.5 machine/m^2 and a packet error rate of 10^{-9} [84]. Associated to this use case, we can include tele-surgery, tele-diagnosis and tele-rehabilitation, where a human control the robot via tactile internet in real-time. Here latency needs to be of 1 ms and reliability of at least 10^{-5} [85].

2.4 Localization

2.4.1 General state of the art

Localization has been part of all the wireless generations from First-Generation (1G) to 5G [86] and can provide location information in particular for emergency or commercial services, advertising, network optimization, mapping, navigation, intelligent transportation, etc. The 1G wireless radio network was introduced based on analog technologies in 1978 [87] where the first location method based on signal strength was used to enhance the communication performance for vehicles [88, 89]. This enables proprietary location solutions in order to introduce emergency services [90]. In order to have network synchronization, GPS appears in Second-Generation (2G) wireless radio network, also called Global System for Mobile communications (GSM) where trilateration methods with synchronized BS were an option [91]. GSM is a digital cellular system that comes up in the 1987s adopting Time Division Multiple Access (TDMA) and FDD and introducing for the first time Short Message Service (SMS) [92]. The first localization activity introduced in the standards for GSM was in 1996 and it concerns location requirements on 911 emergency [93]. Estimation techniques like Time of Arrival (ToA), Angle of Arrival (AoA), cell-ID have been also introduced during the 2G and will be detailed in the rest of the manuscript. UE positioning, ToA and cell-ID were improved within Third-Generation (3G) standardization in [94], [95] and [96] respectively, in the context of Universal Mobile Telecommunications System (UMTS) technology, compatible with Code Division Multiple Access (CDMA) system, FDD and TDD mode. In 4G wireless network, standardization of LTE positioning technology starts in 2008 with 3GPP release 8, for low interference positioning subframe [97]. Observed Time Difference Of Arrival (OTDOA) is a downlink positioning method used in LTE and introduced in 3GPP release 9. It is a multilateration method in which UE measures the ToA of received signals from different BSs. Also, positioning reference signal has been introduced in order to properly allow timing measurements of a UE from a BS signal to improve OTDOA positioning performance [98]. Enhancements for 3GPP LTE positioning are:

MIMO with multiple antenna and beamforming, in particular for indoor positioning [99]; Wireless Local Area Network (WLAN) like Bluetooth, Wifi and UWB [100]; and also D2D aided positioning [101]. While 5G is under standardization, use cases like connected vehicles or surgery are expecting cm accuracy. In Release 16 study item, 3GPP will study how 5G can provide such accuracy for indoor and outdoor use cases including frequencies above 6 GHz [82]. In order to achieve such a positioning accuracy, kind of hybrid technology mixing large bandwidth, multiple sensors and cooperation are under study. Overall accuracy across the wireless radio generations is summarized in Fig. 2.2, where details about what localization method accuracy is achieved under which generation technology are provided in Table III in [86].

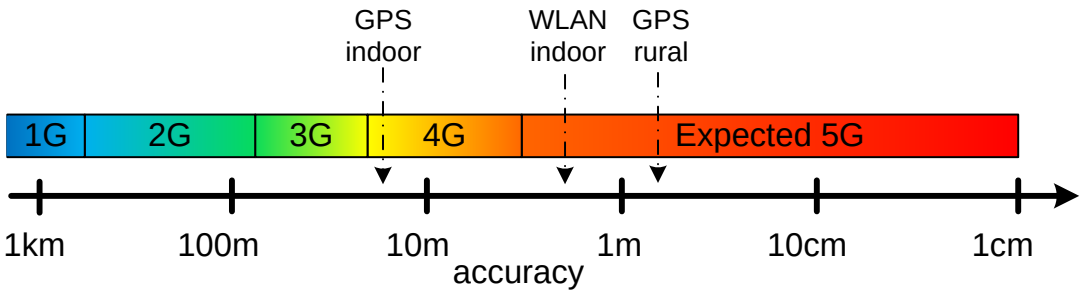


Figure 2.2: 1G to 5G localization accuracy.

2.4.2 Device-to-Device for localization

In the literature, several works are taking advantage of D2D communication for positioning accuracy as it has been studied in wireless sensor networks and UWB [102]. Also [103] proposes to mix localization of target devices and file sharing. A relative positioning method using D2D is introduced. The authors propose to determine devices relative locations by combining devices orientations with acoustic ranging mechanism. The result allows UEs to drag files to destined receivers, and initiate the file sharing in a network group. Authors in [104] propose an adaptive location-sensing system, enabling UEs to estimate their own position cooperating and sharing positioning information though D2D without needing an additional infrastructure. Location-sensing systems use a grid representation that allows the incorporation of external information to improve position estimation accuracy. Regarding distance accuracy, the work in [105] proposes an algorithm reconstructing the Euclidean distance matrix from partially observed distance information providing location map. By casting the low-rank matrix completion problem into the unconstrained minimization problem in Riemannian manifold, authors are able to solve their initial problem by using a modified conjugate gradient algorithm. A good candidate for 5G positioning, mixing cooperative communication and wide bandwidth is

D2D using UWB in unlicensed spectrum. Indeed UWB satisfies *cm* accuracy by emitting short low energy pulses spread in a very large bandwidth. Therefore UWB can also achieve high immunity against interference and fading, low energy consumption [106] and very high accuracy thanks to its wide frequency band as it is indicated by Cramer-Rao Lower Bound (CRLB) [107]. The only considerable disadvantage of UWB is its coverage range, but mixing this technology with D2D routing can extend the coverage. These UWB distinct features offer UWB the best advantages of robustness, accuracy and energy consumption compared to other technologies.

Chapter 3

Device-to-Device aided Multicasting

In this chapter, we consider the multicast channel, where a single transmitter wishes to convey a common message to many receivers in the presence of fading (like in [51, 52, 53]). Since the common message must be decoded by all receivers, the capacity of the fading multicast channel is limited by the worst UE. In the case of the i.i.d. Rayleigh fading channel, it is known that the multicast capacity vanishes inversely to the number n of UEs as n grows [54]. We also consider in this chapter, that the UE receivers are cooperating with each other thanks to D2D technology. We study the D2D-aided multicasting channel without CSIT to overcome the vanishing effect of the multicast rate. We consider a general network topology where the transmitter has only statistical channel knowledge. In this setup, we address the fundamental question: *can D2D without CSIT increase the achievable multicasting rate?*

We propose a two-phase transmission scheme assuming no CSIT: in the first stage the transmitter sends a common message, while in the second stage the subset of UEs who have successfully decoded retransmit this information simultaneously by using the same codeword. We study this two-stage scheme for two metrics of interest:

- The average multicast rate, which is the expected number of successfully decoded bits per channel use achieved by a UE chosen uniformly at random;
- The multicast outage rate, which is the maximum rate at which all the UEs can decode the message with an error probability of ϵ.

We show that by carefully choosing the transmission rate as a function of the network topology we can answer positively to our main question. Namely, our two-phase scheme achieves the following results in the regime of a large number n of UEs: the proposed scheme guarantees an average multicast rate of $\frac{1}{2}\log_2(1+\beta\ln n)$ with high probability for any $\beta < \beta^\star$ where $\beta^\star$ depends on the network topology; a non-vanishing multicast outage rate $\frac{1}{2}\log_2(1+s)$ where s is determined by the target error probability ϵ and

the topology. Moreover, we propose two methods to enhance the proposed two-phase scheme. On one hand, we optimize the time resources between two phases, i.e. instead of considering equal time allocation for both phases, we consider a fraction q for the first (downlink) phase and $1-q$ for the second (D2D) phase. Interestingly, the optimal fraction q becomes arbitrarily closed to 1 in the regime of a large number of UEs, implying that the average multicast rate scaling of $\log_2(1+\beta \ln n)$ is achievable. On the other hand, we incorporate HARQ in order to combine the received signals from the two phases and further decrease the error probability. We consider both Chase Combining (CC) and Incremental Redundancy (IR) coding. The former is simpler and practical, while the latter is the most powerful type of HARQ [108].

In summary, our contributions are outlined below:

1) We propose a simple two-phase scheme that achieves a scalable average multicast rate while requiring only statistical channel knowledge at the transmitter and local channel knowledge at each receiver. The enhanced two-phase scheme with the optimal resource split achieves the average multicast rate of $\log_2(1+\beta \ln n)$ for any $\beta < \beta^\star$ with high probability. This scheme undergoes a phase transition at threshold $\beta^\star \ln n$ where transmissions are successful/unsuccessful with high probability when the SNR is above/below this threshold.

2) We derive tractable asymptotic expressions for both the average and outage multicast rate in the asymptotic regime of a large number of UEs. It is worth noticing that the derivations of such deterministic equivalent expressions are novel and involved based on concentration inequalities.

3) We further provide approximated expressions for both the average and outage multicast rate. Our numerical examples show that these expressions are very tight even for a finite dimension.

A similar two-phase scheme has been studied in [69]. However this work is different from ours in its assumption and concept. First, the channel statistic of all UEs is assumed to be symmetric in [69]. In fact, this is a special case of our model corresponding to a single class $C = 1$ (see Subsection 3.1.4). Second, the metric of the work [69] is on the error probability decay by exploiting channel hardening via space-time code, while we aim at achieving a scalable multicast rate by exploiting multiuser diversity.

This chapter is organized as follows: Section 3.1 presents the system model. Section 3.2 introduces some auxiliary results used for the proofs and also presents the most natural baseline of our scheme. Sections 3.3 and 3.4 studies the average and outage multicast rate metrics respectively. Extensions are analyzed in Section 3.5, firstly regarding resource utilization in Section 3.5.1 and secondly considering the use of HARQ

in Section 3.5.2. Numerical experiments are provided in Section 3.6. Then Section 3.7 concludes the chapter and the appendix is given in Section 3.8 where a list of used parameters is provided.

3.1 Model

3.1.1 Channel Model

We consider D2D-aided multicasting channel assuming that the downlink and the D2D communication share the same bandwidth (inband) and are operated in TDMA. Each node is assumed to be half-duplex.

A BS indexed by 0, wishes to convey a common message to n UEs indexed by $i = 1, \ldots, n$. In each channel use, for both the downlink and D2D, the received signal $y_i[t]$ of UE i can be written as:

$$y_i[t] = \sum_{j=0}^{n} x_j[t] h_{j,i} + N_i[t], \quad t = 1, \ldots, T$$

where T is the total transmission time in channel uses, $x_j[t]$ the signal transmitted by j; $\mathbf{H} = (h_{i,j})_{i,j=0,\ldots,n}$ are the channel coefficients, and $N_i[t] \sim \mathcal{N}(0,1)$ is Additive White Gaussian Noise (AWGN). During the downlink transmission, only the BS transmits the signal, i.e. $x_j[t] = 0$ for $j \neq 0$. During D2D transmission, a given UE i receives the signals from some other UEs (to be specified shortly). The channel coefficients $(h_{i,j})_{i,j=0,\ldots,n}$ are assumed to be independent, Gaussian complex random variables with mean 0 and variance $\gamma_{i,j} = \mathbb{E}(|h_{i,j}|^2)$. The transmitted signals have normalized unit power $\mathbb{E}(|x_i|^2) \leq 1$ for all i. This is without loss of generality because we can always replace x_i by $\frac{x_i}{\sqrt{\mathbb{E}(|x_i|^2)}}$ and $h_{i,j}$ by $\sqrt{\mathbb{E}(|x_i|^2)} h_{i,j}$. The matrix $\mathbf{\Gamma} = (\gamma_{i,j})_{i,j=0,\ldots,n}$, with $\gamma_{i,i} = 0 \ \forall i$, represents the channel statistics and captures the topology of the network. We assume a block-fading channel such that channel coefficients remain constant over a slot of T channel uses and then change from slot to another in an i.i.d. fashion. Under the assumption of statistical CSIT, the transmission strategy should only depend on $\mathbf{\Gamma}$ but not on the channel coefficients $\mathbf{H}$.

3.1.2 Proposed scheme

We consider a two-phase transmission scheme such that time resources T are shared between the downlink and D2D communications. Let $\mathbf{H}$ and $\mathbf{H}'$ denote the channel coefficients of the downlink channel and D2D channels, respectively. For the time-being, we assume that the time resources are equally split between two phases.

- During the first phase of length $\frac{T}{2}$, the BS multicasts a message at rate $\log_2(1+s)$, where s is a parameter of the scheme and can be chosen depending on the channel statistics $\mathbf{\Gamma}$.
- At the end of the first phase, UE i decodes successfully the message if and only if the received SNR is greater than s, i.e. $|h_{0,i}|^2 \geq s$ [1].
- The successfully decoded UEs in the first phase encode and retransmit the message in the second phase [2]. As before, UE i decodes successfully at the end of the second phase if the SNR is above s, i.e. $|\sum_{j=1}^{n} Z_j(s) h'_{j,i}|^2 \geq s$, where $Z_1(s), \ldots, Z_n(s)$ are binary variables with $Z_j(s) = 1$ if j has decoded successfully during the first phase and 0 otherwise.

Notice that the second phase of our scheme is different from that considered in [69]. Namely, the SNR of UE i is non-coherently combined in our case while it is coherently combined streams in [69].

3.1.3 Performance Measures

We say that UE i decodes if and only if it successfully decodes either in the first or second phase. By letting $P_i(s)$ denote the *individual success probability* of UE $i = 1, \ldots, n$, we define two other related probabilities:

- $\bar{P}(s) \stackrel{\Delta}{=} \frac{1}{n} \sum_{i=1}^{n} P_i(s)$ denotes the *average success probability* of a UE chosen uniformly at random amongst $i = 1, \ldots, n$.
- $P_+(s)$ denotes the *joint success probability*, i.e. the probability that the all UEs decode successfully.

We study two performance measures for our problems. We define the average multicast rate as

$$R^m = \frac{1}{2} \max_{s \geq 0} \{\log_2(1+s) \bar{P}(s)\}, \tag{3.1}$$

which is the expected number of bits per channel use received by a UE chosen uniformly at random. Note that the pre-log factor of $\frac{1}{2}$ captures the resource loss due to the D2D phase where no new information is conveyed.

We define the multicast outage rate as

$$R^o = \frac{1}{2} \log_2(1+s) \text{ with } s \text{ solution to } P_+(s) = 1 - \epsilon, \tag{3.2}$$

[1] This is valid by assuming that T is arbitrarily large so that one may use a capacity-achieving code.

[2] Here we assume that the codebook is shared by all nodes a priori so that each UE can encode the decoded message.

which is the largest rate such that *all* UEs jointly decode with probability at least $1-\epsilon$, with ϵ some fixed reliability level. Both performance measures are relevant to scenarios of interest. On one hand, R^m is relevant to delay-tolerant application such as video streaming, whose goal is to ensure that most UEs receive enough information on average. On the other hand, the multicast outage rate R^o appears more suited to safety applications, where *all* UEs are guaranteed to decode the message with a desired reliability.

3.1.4 Block Model

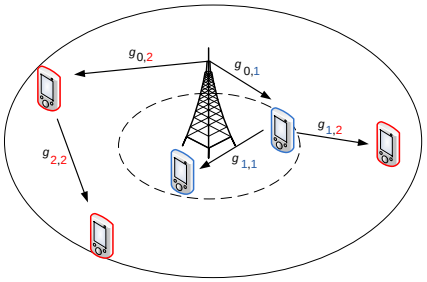


Figure 3.1: Block Model.

To study large systems in a tractable manner, we introduce a block structure to the matrix $\mathbf{\Gamma}$. Users $1,\dots,n$ are partitioned in classes $1,\dots,C$, where $c_i \in \{1,\dots,C\}$ indicates the class of UE i, and $c_0 = 0$ by convention. There are $n\alpha_c$ UEs of class c, with $\boldsymbol{\alpha} = (\alpha_c)_{c=0,\dots,C}$ and $\alpha_c > 0$ the proportion of UEs of class c and $\sum_{c=1}^{C} \alpha_c = 1$. Matrix $\mathbf{\Gamma}$ is a block matrix, so that the mean channel gains between two UEs depends solely on their class. Namely, $\gamma_{i,j} = g_{c_i,c_j}$ for all $i,j = 0,\dots,n$. Define $G = \max_{c,c'=0,\dots,C} g_{c,c'}$ the largest entry of $\mathbf{\Gamma}$, so that $\max_{i,j=0,\dots,n} \gamma_{i,j} = G$. This model can represent any network topology, as the number of classes may be arbitrary.

Example 3.1.1. *Single class ($C = 1$): A single class with $\alpha = (1)$ represents an homogeneous network where the downlink channel gain $g_{0,1}$ is equal for all the UEs and the D2D channel gain $g_{1,1}$ between the UEs is also equal.*

Example 3.1.2. *Two classes ($C = 2$): This case represents a network of two clusters with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ such that cluster i has $\alpha_i n$ UEs and is illustrated in Fig. 3.1. Class 1 and 2 are represented in blue and red respectively. The downlink channel gain is equal for the UEs in the same cluster, e.g. $g_{0,1}$ for UEs in cluster 1 and $g_{0,2}$ for UEs in 2. The same stands for the D2D channel gain between the UEs inside the same cluster, $g_{1,1}$ and $g_{2,2}$ respectively. The D2D channel gain across clusters is equal: $g_{1,2} = g_{2,1}$.*

We introduce the following natural assumption.

Assumption 3.1.1. *The BS can communicate with any class of UEs in two hop transmissions so that for all $c = 1, \dots, C$ there exists c' such that $g_{0,c'} g_{c',c} > 0$.*

3.2 Preliminaries

3.2.1 Auxiliary results

We recall a few basic concentration inequalities. Proofs may be found in [109].

Proposition 3.2.1. *[Chebychev's inequality] Consider Y a random variable. Then for all $\delta > 0$ we have $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \delta) \leq \frac{var(Y)}{\delta^2}$.*

Proposition 3.2.2. *[Hoeffding's inequality] Consider $Y_1, \dots, Y_n$ independent random variables in $[0,1]$. Then for all $\delta > 0$ we have $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)) \geq \delta\right) \leq e^{-2n\delta^2}$.*

Proposition 3.2.3. *[Chernoff's inequality] Consider $Y_1, \dots, Y_n$ independent random variables in $[0,1]$. Define $y_n = \sum_{i=1}^n \mathbb{E}(Y_i)$. Then for all $\delta \in [0,1]$ we have $\mathbb{P}\left(\sum_{i=1}^n Y_i \geq (1+\delta) y_n\right) \leq e^{-\frac{y_n \delta^2}{3}}$ and $\mathbb{P}\left(\sum_{i=1}^n Y_i \leq (1-\delta) y_n\right) \leq e^{-\frac{y_n \delta^2}{3}}$.*

3.2.2 Asymptotique Notations

Let $n \in \mathbb{N}$, f and g both real functions.
$f(n) \underset{n\to\infty}{=} \mathcal{O}(g(n))$ means that $\limsup_{n\to\infty} \frac{f(n)}{g(n)} \leq \infty$; $f(n) \underset{n\to\infty}{\sim} g(n)$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

3.2.3 Baseline

We will compare the performance of our proposed scheme to the baseline without D2D. Namely, the BS multicasts a message at rate $\log_2(1+s)$ at each time slot. The performance of the baseline is recalled below.

Proposition 3.2.4. *The performance of the baseline scheme is:*

$$R^m = \max_{s \geq 0} \{\log_2(1+s) \sum_{c=1}^{C} \alpha_c e^{-\frac{s}{g_{0,c}}}\} \underset{n\to\infty}{=} \mathcal{O}(1),$$

$$R^o = \log_2\Big(1 + \frac{1}{n}\ln\Big(\frac{1}{1-\epsilon}\Big)\Big(\sum_{c=1}^{C} \frac{\alpha_c}{g_{0,c}}\Big)^{-1}\Big) \underset{n\to\infty}{=} \mathcal{O}\Big(\frac{1}{n}\Big).$$

Proof. In the baseline scheme, UE i successfully decodes if and only if $|h_{0,i}|^2 \geq s$. Since $|h_{0,i}|^2$ follows an exponential distribution with mean $\gamma_{0,i}$, we have the *individual success probability*:

$$P_i(s) = \mathbb{P}(|h_{0,i}|^2 \geq s) = e^{-\frac{s}{\gamma_{0,i}}},$$

and the *average success probability*:

$$\bar{P}(s) = \frac{1}{n}\sum_{i=1}^{n} e^{-\frac{s}{\gamma_{0,i}}} = \sum_{c=1}^{C} \alpha_c e^{-\frac{s}{g_{0,c}}}.$$

By plugging the above expression into (3.1), we obtain the desired average multicast rate.

Also, the transmission is *jointly successful* if and only if $|h_{0,i}|^2 \geq s$ for all i. Recalling that $|h_{0,i}|^2$ follows an exponential distribution with mean $\gamma_{0,i}$ and that $h_{0,1}, \ldots, h_{0,n}$ are independent, the *joint success probability* is:

$$\begin{aligned} P_+(s) &= \mathbb{P}(|h_{0,i}|^2 \geq s\,,\ i = 1, \ldots, n) = \prod_{i=1}^{n} \mathbb{P}(|h_{0,i}|^2 \geq s) \\ &= \exp\left\{-s\sum_{i=1}^{n}\frac{1}{\gamma_{0,i}}\right\} = \exp\left\{-sn\sum_{c=1}^{C}\frac{\alpha_c}{g_{0,c}}\right\}. \end{aligned}$$

The multicast outage rate is $\log_2(1+s)$ with s solution to the equation $P_+(s) = 1-\epsilon$, hence finding s and replacing in (3.2) yields the announced result. □

3.3 Average Multicast Rate

In this section we study the average multicast rate of the proposed two-phase scheme. Interestingly, the scheme undergoes a phase transition in the regime of a large number of UEs $n \to \infty$, so that the *average success probability* $\bar{P}(s)$ becomes constant-by-parts and may be computed explicitly as a function of the network topology, i.e. Γ and $\boldsymbol{\alpha}$.

3.3.1 Individual success probability

We first provide the *individual success probability* $P_i(s)$ expression in Proposition 3.3.1, which will serve as the backbone of our analysis. To this end, we define $Z_1(s), \ldots, Z_n(s)$ as independent random variables in $\{0, 1\}$ and

$$X_i(s) \triangleq \sum_{j=1; j\neq i}^{n} Z_j(s)\gamma_{j,i} \tag{3.3}$$

Proposition 3.3.1. *For any $i = 1, \ldots, n$ and $s \geq 0$ we have:*

$$P_i(s) = 1 - (1 - e^{-\frac{s}{\gamma_{0,i}}})\mathbb{E}\left(1 - \exp\left\{-\frac{s}{X_i(s)}\right\}\right).$$

This result shows that the *individual success probability* $P_i(s)$ can be controlled by examining the fluctuations of $X_i(s)$, a sum of independent Bernoulli random variables.

Proof. Consider $\mathbf{H} = (h_{i,j})_{i,j=0,\ldots,n}$ and $\mathbf{H}' = (h'_{i,j})_{i,j=0,\ldots,n}$ the channel coefficients during the first and second time slot respectively. By assumption $\mathbf{H}$ is independent from $\mathbf{H}'$. We define $Z_i(s) \stackrel{\Delta}{=} \mathbb{1}\{|h_{0,i}|^2 \geq s\}$, where $Z_i(s) = 1$ if UE i decodes correctly at the first time slot, and 0 otherwise. Also, denote by $\mathbf{Z}(s) = (Z_1(s), \ldots, Z_n(s))$ the outcome of the first time slot. Since $|h_{0,i}|^2$ has exponential distribution with mean $\gamma_{0,i}$, it follows that

$$\mathbb{E}(Z_i(s)) = e^{-\frac{s}{\gamma_{0,i}}}. \tag{3.4}$$

Furthermore, since $h_{0,1}, \ldots, h_{0,n}$ are independent, so are $Z_1(s), \ldots, Z_n(s)$.

Consider UE i. Conditionally to the value of $\mathbf{Z}(s)$, UE i does not decode successfully in the first phase if and only if $Z_i(s) = 0$. If $Z_i(s) = 0$, he does not decode successfully in the second phase if and only if $|\sum_{j\neq i} h'_{j,i} Z_j(s)|^2 \leq s$, where $|\sum_{j\neq i} h'_{j,i} Z_j(s)|^2$ has exponential distribution with mean $\sum_{j\neq i}^{n} Z_j(s)\gamma_{j,i} = X_i(s)$, since $\mathbf{Z}(s)$ is independent of $\mathbf{H}'$. Hence

$$\mathbb{P}(i \text{ does not decode}|\mathbf{Z}(s)) = (1 - Z_i(s))(1 - e^{-\frac{s}{X_i(s)}}).$$

Taking the expectation over $\mathbf{Z}(s)$, we have

$$\begin{aligned}\mathbb{E}[\mathbb{P}(i \text{ does not decode}|\mathbf{Z}(s))] &= \mathbb{P}(Z_i(s) = 0)\mathbb{E}(1 - e^{-\frac{s}{X_i(s)}}) \\ &= (1 - e^{-\frac{s}{\gamma_{0,i}}})\mathbb{E}(1 - e^{-\frac{s}{X_i(s)}}).\end{aligned}$$

This yields the desired individual success probability. □.

3.3.2 Asymptotic behavior

We now analyze how the average multicast rate scales in the regime of a large number of UEs. We define

$$\beta_c = \max_{c'=1,\ldots,C}\{g_{0,c'}\,\mathbb{1}\{g_{c',c} > 0\}\},$$

as the largest value of $g_{0,c'}$ such that class c' can communicate with class c in the second phase i.e. $g_{c',c} > 0$. Further define the minimum value $\beta^\star = \min_{c=1,\ldots,C} \beta_c$.

Theorem 3.3.1.

(i) For any $\beta > 0$ and $i = 1, \ldots, n$ we have the individual success probability:

$$P_i\left(\beta \ln n\right) \underset{n\to\infty}{\to} \begin{cases} 1 & \text{if } \beta < \beta_{c_i} \\ 0 & \text{otherwise.} \end{cases}$$

(ii) and the average success probability:

$$\bar{P}\left(\beta \ln n\right) \underset{n\to\infty}{\to} \sum_{c=1}^{C} \alpha_c \mathbb{1}\{\beta < \beta_c\},$$

Theorem 3.3.1 shows that in the limit of a large number of UEs the *individual success probability* P_i undergoes a phase transition at the value $\beta_{c_i} \ln n$ for any UE i. Namely transmissions are always successful above this threshold and unsuccessful below it.
Three remarks are in order:

- Our scheme transmits at average multicast rate $\frac{1}{2}\log_2(1+\beta \ln n)$ with an arbitrarily high probability of success for any $\beta < \beta^\star$.
- As $n \to \infty$, the average multicast rate of our scheme scales as $\mathcal{O}(\ln \ln n)$ while the baseline yields an average multicast rate of $\mathcal{O}(1)$. So considering two slots instead of one yields a dramatic impact on performance. Actually, as n and s increase, only a small part of the UEs will be able to decode in a common multicast scheme. Yet, the choice of β ensures that the number of decoded UEs will increase unboundedly in our scheme, and these UEs are responsible for the transmission to other UEs.
- To obtain an order-optimal rate, one can choose $s = \beta \ln n$ for β slightly smaller than $\beta^\star$, so that optimizing over s is not needed.

Proof. From Proposition 3.3.1, we have:

$$P_i(\beta \ln n) = 1 - (1 - n^{-\frac{\beta}{g_{0,c_i}}})\mathbb{E}(1 - e^{-\frac{\beta \ln n}{X_i(\beta \ln n)}}).$$

We have $n^{-\frac{\beta}{g_{0,c_i}}} \underset{n\to\infty}{\to} 0$ so that:

$$\lim_{n\to\infty} P_i(\beta \ln n) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{\beta \ln n}{X_i(\beta \ln n)}}).$$

The $\beta > \beta_{c_i}$ case. Assume that $\beta > \beta_{c_i}$, and we control the expectation of $X_i(\beta \ln n)$. Since, from (3.4)
$\mathbb{E}(Z_j(\beta \ln n)) = e^{-\frac{\beta \ln n}{\gamma_{0,j}}} = n^{-\frac{\beta}{\gamma_{0,j}}}$ and from Section 3.1.4:

$$\begin{aligned}\mathbb{E}\Big(\frac{X_i(\beta \ln n)}{\beta \ln n}\Big) &= \sum_{j\neq i} \frac{\gamma_{j,i} n^{-\frac{\beta}{\gamma_{0,j}}}}{\beta \ln n} \leq \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}{\beta \ln n} \\ &\leq \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{\beta_{c_i}}}}{\beta \ln n} \underset{n\to\infty}{\to} 0\end{aligned}$$

since for all c either $g_{c,c_i} = 0$ or $g_{0,c} \leq \beta_{c_i}$, and $\beta > \beta_{c_i}$. Therefore, $\frac{X_i(\beta \ln n)}{\beta \ln n}$ converges to 0 in L^1 so that it converges to 0 in distribution as well. Since $x \mapsto e^{-\frac{1}{x}}$ is both continuous and bounded we get:

$$\lim_{n\to\infty} P_i(\beta \ln n) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{\beta \ln n}{X_i(\beta \ln n)}}) = 0.$$

Hence the *individual success probability* P_i is null in the case of $\beta > \beta_{c_i}$ when the number of UEs is large ($n \to \infty$). As a consequence the *average success probability* $\bar{P}$ in the same regime is also null.

The $\beta < \beta_{c_i}$ case. Now consider $\beta < \beta_{c_i}$. We control the moments of $X_j(s)$, since from (3.3) and (3.4):

$$\mathbb{E}(X_i(s)) = \sum_{j\neq i} \gamma_{j,i} e^{-\frac{s}{\gamma_{0,j}}},$$

and since $Z_1(s), \ldots, Z_n(s)$ are independent:

$$\begin{aligned}\mathrm{var}(X_i(s)) &= \sum_{j\neq i} \gamma_{j,i}^2 e^{-\frac{s}{\gamma_{0,j}}}(1 - e^{-\frac{s}{\gamma_{0,j}}}) \\ &\leq G \sum_{j\neq i}^{n} \gamma_{j,i} e^{-\frac{s}{\gamma_{0,j}}} = G\mathbb{E}(X_i(s)).\end{aligned}$$

where G is a constant value.
Apply Chebychev's inequality:

$$\begin{aligned}\mathbb{P}\left(X_i(s) \leq \frac{\mathbb{E}(X_i(s))}{2}\right) &\leq \mathbb{P}\left(|X_i(s) - \mathbb{E}(X_i(s))| \geq \frac{\mathbb{E}(X_i(s))}{2}\right) \\ &\leq \frac{4\mathrm{var}(X_i(s))}{\mathbb{E}(X_i(s))^2} \leq \frac{4G}{\mathbb{E}(X_i(s))}\end{aligned}$$

using the previous bound. Hence

$$\mathbb{P}\left(X_i(s) \geq \frac{\mathbb{E}(X_i(s))}{2}\right) \geq 1 - \frac{4G}{\mathbb{E}(X_i(s))}. \tag{3.5}$$

Since $x \mapsto e^{-\frac{1}{x}}$ is increasing:

$$\mathbb{E}(e^{-\frac{\beta \ln n}{X_i(\beta \ln n)}}) \geq \mathbb{P}\left(X_i(\beta \ln n) \geq \frac{\mathbb{E}(X_i(\beta \ln n))}{2}\right) e^{-\frac{2\beta \ln n}{\mathbb{E}(X_i(\beta \ln n))}}.$$

Consider $\hat{c}$ such that $g_{\hat{c},c_i} > 0$ and $g_{0,\hat{c}} = \beta_{c_i}$. Then:

$$\begin{aligned}\mathbb{E}\left(\frac{X_i(\beta \ln n)}{\beta \ln n}\right) &= -\frac{\gamma_{ii} n^{-\frac{\beta}{\gamma_{0,i}}}}{\beta \ln n} + \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}{\beta \ln n} \\ &\geq -\frac{\gamma_{ii} n^{-\frac{\beta}{\gamma_{0,i}}}}{\beta \ln n} + \frac{\alpha_{\hat{c}} g_{\hat{c},c_i} n^{1-\frac{\beta}{\beta_{c_i}}}}{\beta \ln n} \underset{n\to\infty}{\to} \infty.\end{aligned}$$

Replacing in (3.5) we deduce

$$\mathbb{P}\left(X_i(\beta \ln n) \geq \frac{\mathbb{E}(X_i(\beta \ln n))}{2}\right) \underset{n\to\infty}{\to} 1,$$

and $e^{-\frac{2\beta \ln n}{\mathbb{E}(X_i(\beta \ln n))}} \underset{n\to\infty}{\to} 1$ so that:

$$\lim_{n\to\infty} P_i(\beta \ln n) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{\beta \ln n}{X_i(\beta \ln n)}}) = 1.$$

which completes the proof of statement (i). Hence, in this case and in the regime of large number of UEs, *the individual success probability* of decoding is closed to 1. As a consequence, this clearly highlight a transition phase at value β_{c_i}, where this *individual success probability* becomes constant by parts (1 when $\beta < \beta_{c_i}$ and 0 $\beta > \beta_{c_i}$). Statement (ii) follows from the fact that *average success probability* $\bar{P}(s) = \frac{1}{n}\sum_{i=1}^{n} P_i(s)$.

□

Corollary 3.3.2. *For any $\beta < \beta^\star$ we have:*

$$R^m \geq \frac{1 - o(1)}{2} \log_2(1 + \beta \ln n)\,,\ n \to \infty.$$

3.3.3 Non asymptotic behavior

We now state Theorem 3.3.3, a further result which gives a tractable, accurate approximation for $1-P_i(\beta \ln n)$ in the regime where $P_i(\beta \ln n) \underset{n\to\infty}{\to} 1$, i.e. whenever $\beta < \beta_{c_i}$.
Two main facts are worth mentioning:

- This approximation is very accurate even for modest size systems (say $n \geq 50$) as shown by our numerical experiments in Section 3.6.
- Due to its accuracy, it allows to find the optimal value of s given $(g_{c_i,c_j})_{c_i,c_j=0,\ldots,C}$ and $\boldsymbol{\alpha}$ in a tractable manner, so that finite size systems can be dealt efficiently.

Theorem 3.3.3. *Consider $\beta < \beta_c$, then:*

$$1 - P_i\left(\beta \ln n\right) \underset{n\to\infty}{\sim} 1 - \exp\left\{\frac{-\beta \ln n}{\sum_{c=1}^{C} \alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}\right\}.$$

Proof. Define $f(x) = 1 - e^{-\frac{1}{x}}$. Throughout the proof consider i fixed, and define $V = \frac{X_i(\beta \ln n)}{\beta \ln n}$.

From Proposition 3.3.1:

$$1 - P_i(\beta \ln n) = (1 - n^{-\frac{\beta}{\gamma_{0,i}}})\mathbb{E}(f(V)) \underset{n\to\infty}{\sim} \mathbb{E}(f(V)).$$

since $n^{-\frac{\beta}{\gamma_{0,i}}} \underset{n\to\infty}{\to} 0$. Define:

$$v_n = \mathbb{E}(V) = -\frac{\gamma_{ii} n^{-\frac{\beta}{\gamma_{0,i}}}}{\beta \ln n} + \frac{\sum_{c=1}^{C} \alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}{\beta \ln n} \underset{n\to\infty}{\to} \infty.$$

since $\beta < \beta^\star$ and recalling that $\gamma_{ii} = 0 \forall i$. To complete the proof, it suffices to show that

$$\mathbb{E}(f(V)) \underset{n\to\infty}{\sim} f(v_n).$$

Now consider $\delta \in [0,1]$ fixed and let us bound $\mathbb{E}(f(V))$.

Upper bound: Since $x \mapsto f(x)$ is decreasing:

$$\begin{aligned} f(V) &= f(V)\mathbb{1}\{V \leq (1-\delta)v_n\} + f(V)\mathbb{1}\{V \geq (1-\delta)v_n\} \\ &\leq \mathbb{1}\{V \leq (1-\delta)v_n\} + f((1-\delta)v_n). \end{aligned}$$

So taking expectations and using Chernoff's inequality (recalled in proposition 3.2.3 in Section 3.2):

$$\begin{aligned} \mathbb{E}(f(V)) &\leq \mathbb{P}(V \leq (1-\delta)v_n) + f((1-\delta)v_n) \\ &\leq e^{-\frac{\delta^2 v_n}{3}} + f((1-\delta)v_n). \end{aligned}$$

Now using the facts $v_n \underset{n\to\infty}{\to} \infty$ so that $f(v_n) \underset{n\to\infty}{\sim} \frac{1}{v_n}$ and $v_n e^{-\frac{\delta^2 v_n}{3}} \underset{n\to\infty}{\to} 0$ we get:

$$\limsup_{n\to\infty} \frac{\mathbb{E}(f(V))}{f(v_n)} \leq \frac{1}{1-\delta}.$$

Lower bound: Similarly:

$$\begin{aligned} f(V) &= f(V)\mathbb{1}\{V \leq (1+\delta)v_n\} + f(V)\mathbb{1}\{V \geq (1+\delta)v_n\} \\ &\geq f((1+\delta)v_n)\mathbb{1}\{V \leq (1+\delta)v_n\} \end{aligned}$$

So taking expectations and using Chernoff's inequality (recalled in proposition 3.2.3 in Section 3.2):

$$\begin{aligned} \mathbb{E}(f(V)) &\geq f((1+\delta)v_n)\mathbb{P}(V \leq (1+\delta)v_n) \\ &\geq f((1+\delta)v_n)(1 - e^{-\frac{\delta^2 v_n}{3}}). \end{aligned}$$

Using the facts $v_n \underset{n\to\infty}{\to} \infty$ so that $f(v_n) \underset{n\to\infty}{\sim} \frac{1}{v_n}$ and $e^{-\frac{\delta^2 v_n}{3}} \underset{n\to\infty}{\to} 0$ we get:

$$\liminf_{n\to\infty} \frac{\mathbb{E}(f(V))}{f(v_n)} \geq \frac{1}{1+\delta}.$$

Putting it together: The above holds for any $\delta \in (0,1)$. So we have proven that δ arbitrarily small:

$$\frac{1}{1+\delta} \leq \liminf_{n\to\infty} \frac{\mathbb{E}(f(V))}{f(v_n)} \leq \limsup_{n\to\infty} \frac{\mathbb{E}(f(V))}{f(v_n)} \leq \frac{1}{1-\delta}.$$

Hence we have proven

$$\mathbb{E}(f(V)) \underset{n\to\infty}{\sim} f(v_n).$$

which concludes the proof. □

3.4 Multicast Outage Rate

We now study the multicast outage rate and show that it can be computed explicitly as a function of the network topology ($\mathbf{\Gamma}$ and $\boldsymbol{\alpha}$) in the regime of a large number of UEs. We further show that, while the outage rate of the baseline scheme vanishes when $n \to \infty$, our scheme guarantees a constant multicast outage rate.

3.4.1 Joint success probability

As in the average multicast case, we express the multicast outage rate as a function of $X_1(s), \ldots, X_n(s)$ in Proposition 3.4.1.

Proposition 3.4.1. *For any $s \geq 0$ we express the joint success probability as:*

$$P_+(s) = \mathbb{E}\left[\exp\left\{-s\sum_{i=1}^{n} \frac{1-Z_i(s)}{X_i(s)}\right\}\right],$$

where $Z_1(s), \ldots, Z_n(s)$ are independent random variables in $\{0,1\}$ and $\mathbb{E}(Z_i(s))$, $X_i(s)$ are defined for $i = 1, \ldots, n$ in (3.4) *and* (3.3) *respectively.*

Proof. We use the same notation as in the proof of Proposition 3.3.1. User i successfully decodes if and only if either the first phase or the second phase is

successful, $Z_i(s) = 1$, or $|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}|^2 \geq s$ respectively. Combining these two, UE i decodes if and only if:

$$\left| \sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i} \right|^2 \geq s(1 - Z_i(s)).$$

Now, considering all the UEs:

$$\begin{aligned}
\mathbb{P}(\text{all decode}|\mathbf{Z}(s)) &= \mathbb{P}\Big(\Big| \sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i} \Big|^2 \geq s(1 - Z_i(s)) \,,\, \forall i|\mathbf{Z}(s) \Big) \\
&= \prod_{i=1}^{n} \mathbb{P}\Big(\Big| \sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i} \Big|^2 \geq s(1 - Z_i(s))|\mathbf{Z}(s) \Big) \\
&= \exp\left\{ -s \sum_{i=1}^{n} \frac{1 - Z_i(s)}{X_i(s)} \right\}.
\end{aligned}$$

where the last inequality follows conditional to $\mathbf{Z}(s)$, also $|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}|^2$ that has exponential distribution with mean $X_i(s)$, and due to the independent entries of $\mathbf{H}'$. Averaging over $\mathbf{Z}(s)$ yields the result. □

3.4.2 Asymptotic behavior

We now prove our main result concerning multicast outage rate stated in Theorem 3.4.1. The main proof element is to show that, for any s random variable $\sum_{i=1}^{n} \frac{1-Z_i(s)}{X_i(s)}$ concentrates around its expectation when the number of UEs n grows large. The following facts should be noted:

- In the regime of a large number of UEs, the multicast outage rate of our proposed two-phase scheme converges to a non-zero finite value. Recall that the baseline scheme only guarantees a vanishing multicast outage rate. Hence our scheme is very efficient in countering the variability of the channel coefficients and induces a perfect channel hardening.

- The multicast outage rate can be computed explicitly as a function of $\boldsymbol{\alpha}$ and $(g_{c_i,c_j})_{c_i,c_j=0,\ldots,C}$ as we will see in corollary 3.4.2 by a zero-finding method such as bisection or Newton-Raphson.

- In fact, this asymptotic result provides a very good approximation even for systems of modest size (say $n \geq 10$), as shown by numerical experiments (see Section 3.6).

Theorem 3.4.1. *For any $s \geq 0$ we express the joint success probability as:*

$$P_+(s) \underset{n\to\infty}{\to} \exp\left\{-s\sum_{c=1}^{C}\frac{\alpha_c(1-e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C}\alpha_{c'}g_{c',c}e^{-\frac{s}{g_{0,c'}}}}\right\}.$$

Proof. We first prove the following fact:

$$\sum_{i=1}^{n}\frac{1-Z_i(s)}{X_i(s)} \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} \sum_{i=1}^{n}\frac{1-e^{-\frac{s}{\gamma_{0,i}}}}{\mathbb{E}(X_i(s))}.$$

We bound the error as follows:

$$\left|\sum_{i=1}^{n}\frac{1-Z_i(s)}{X_i(s)}-\sum_{i=1}^{n}\frac{1-e^{-\frac{s}{\gamma_{0,i}}}}{\mathbb{E}(X_i(s))}\right| \leq \left|\sum_{i=1}^{n}\frac{1-Z_i(s)}{X_i(s)}-\frac{1-Z_i(s)}{\mathbb{E}(X_i(s))}\right|+\left|\sum_{i=1}^{n}\frac{Z_i(s)-e^{-\frac{s}{\gamma_{0,i}}}}{\mathbb{E}(X_i(s))}\right|.$$

We now prove that both terms on the right hand side (r.h.s.) go to 0 in probability.

First term We have:

$$\begin{aligned}\left|\sum_{i=1}^{n}\frac{1-Z_i(s)}{X_i(s)}-\frac{1-Z_i(s)}{\mathbb{E}(X_i(s))}\right| &\leq \sum_{i=1}^{n}\left|\frac{1}{X_i(s)}-\frac{1}{\mathbb{E}(X_i(s))}\right| \\ &= \sum_{i=1}^{n}\frac{|X_i(s)-\mathbb{E}(X_i(s))|}{X_i(s)\mathbb{E}(X_i(s))}.\end{aligned}$$

Furthermore

$$\mathbb{E}(X_i(s)) = -\gamma_{i,i}e^{-\frac{s}{\gamma_{0,i}}}+n\sum_{c=1}^{C}\alpha_c g_{c,c_i}e^{-\frac{s}{g_{0,c}}} \geq m(s)n$$

recalling that $\gamma_{i,i}=0\forall i$, and with

$$m(s) = \min_{c'=1,\dots,C}\max_{c=1,\dots,C}\left\{\alpha_c g_{c,c'}e^{-\frac{s}{g_{0,c}}}\right\} > 0.$$

where $m(s)>0$ from assumption 3.1.1.

For $i=1,\dots,n$, as we can see in (3.3) $X_i(s)$ is a sum of $n-1$ random variables bounded by G, so that Hoeffding's inequality yields:

$$\mathbb{P}(|X_i(s)-\mathbb{E}(X_i(s))| \geq G\sqrt{n\ln n}) \leq \frac{1}{n^2}.$$

Using a union bound over $i=1,\dots,n$:

$$\begin{aligned}&\mathbb{P}(\max_{i=1,\dots,n}|X_i(s)-\mathbb{E}(X_i(s))| \geq G\sqrt{n\ln n}) \\ &\leq \sum_{i=1}^{n}\mathbb{P}(|X_i(s)-\mathbb{E}(X_i(s))| \geq G\sqrt{n\ln n}) \leq \frac{1}{n}\underset{n\to\infty}{\to}0.\end{aligned}$$

Therefore the event

$$\mathcal{A} = \{\max_{i=1,\dots,n} |X_i(s) - \mathbb{E}(X_i(s))| \leq G\sqrt{n \ln n}\}$$

occurs with high probability when $n \to \infty$. Consider n large enough so that $m(s)n - G\sqrt{n \ln n} \geq 0$. If $\mathcal{A}$ occurs:

$$\sum_{i=1}^{n} \frac{|X_i(s) - \mathbb{E}(X_i(s))|}{X_i(s)\mathbb{E}(X_i(s))} \leq \frac{G\sqrt{n \ln n}}{m(s)n(m(s)n - G\sqrt{n \ln n})} \underset{n\to\infty}{\to} 0.$$

Second term Compute the second moment of the second term:

$$\mathbb{E}\left[\Big(\sum_{i=1}^{n} \frac{Z_i(s) - e^{-\frac{s}{\gamma_{0,i}}}}{\mathbb{E}(X_i(s))}\Big)^2\right] = \sum_{i=1}^{n} \frac{e^{-\frac{s}{\gamma_{0,i}}}(1 - e^{-\frac{s}{\gamma_{0,i}}})}{\mathbb{E}(X_i(s))^2} \leq \sum_{i=1}^{n} \frac{1}{\mathbb{E}(X_i(s))^2} \leq \frac{n}{(m(s)n)^2} \underset{n\to\infty}{\to} 0,$$

since from (3.4) $\mathbb{E}(Z_i(s)) = e^{-\frac{s}{\gamma_{0,i}}}$ and $Z_1(s), \dots, Z_n(s)$ are independent. So the second term goes to 0 in probability, since L^2 convergence implies convergence in probability.

Putting it together we have proven that:

$$\sum_{i=1}^{n} \frac{1 - Z_i(s)}{X_i(s)} \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} \sum_{i=1}^{n} \frac{1 - e^{-\frac{s}{\gamma_{0,i}}}}{\mathbb{E}(X_i(s))} = \sum_{c=1}^{C} \frac{\alpha_c(1 - e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C} \alpha_{c'} g_{c',c} e^{-\frac{s}{g_{0,c'}}}}.$$

Since convergence in probability implies convergence in distribution and $x \mapsto e^{-x}$ is continuous and bounded on $\mathbb{R}_+$ we obtain the result:

$$P_+(s) \underset{n\to\infty}{\to} \exp\left\{-s \sum_{c=1}^{C} \frac{\alpha_c(1 - e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C} \alpha_{c'} g_{c',c} e^{-\frac{s}{g_{0,c'}}}}\right\}.$$

□

Now that the *joint success probability* has been expressed in the regime of large number of UEs, the multicast outage rate can be deduced.

Corollary 3.4.2. *When $n \to \infty$, the multicast outage rate converges to $\frac{1}{2}\log_2(1 + s)$ where s is the unique solution to:*

$$s \sum_{c=1}^{C} \frac{\alpha_c(1 - e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C} \alpha_{c'} g_{c',c} e^{-\frac{s}{g_{0,c'}}}} = \ln\Big(\frac{1}{1 - \epsilon}\Big).$$

and for $\epsilon \to 0$ we have $s \to 0$ and Taylor development gives:

$$s \underset{\epsilon \to 0}{\to} \sqrt{\frac{\ln\left(\frac{1}{1-\epsilon}\right)}{\sum_{c=1}^{C} \frac{\alpha_c}{g_{0,c}} \left(\sum_{c'=1}^{C} \alpha_{c'} g_{c',c} \right)^{-1}}}$$

3.5 Extensions

In this section, we enhance the proposed D2D-aided multicasting scheme in two different ways. First, we examine the case where one is allowed to optimize resource utilization across the two phases of the scheme: instead of considering equal time allocation for two phases, one allocates a fraction q and $1-q$ for the first and the second phase respectively where $q \in [0,1]$ is a parameter of the scheme. Second, we consider the case where the received signals from the two phases are combined using a form of HARQ in order to further decrease the error probability. We consider both Chase Combining (CC) and Incremental Redundancy (IR) coding. We derive two conclusions from these extensions.

- First, the proper resource allocation allows doubling the average multicast rate compared to the time splitting between two phases in the regime of a large number of UEs $n \to \infty$. To achieve this performance gain, it suffices to select q arbitrarily close to 1.
- Although they do not increase the scaling gain, both versions of HARQ yield a non-negligible gain in a finite number of UEs. The IR coding provides a better performance at the price of higher complexity.

3.5.1 Resource Allocation Optimization

We first examine the case where one is allowed to optimize resource utilization across the two phases of our scheme. In this setting our scheme works as follows:

- During the first (downlink) phase of duration qT channel uses, the BS multicasts a message at rate $\log_2(1+s)$, where both $s \geq 0$ and $q \in [0,1]$ are parameters of the scheme and can be chosen depending on the channel statistics $\mathbf{\Gamma}$.
- At the end of the first time slot, UE i decodes successfully if and only if:

  $$q \log_2(1+|h_{0,i}|^2) \geq q \log_2(1+s),$$

  equivalently, if and only if that the received SNR is greater than s, i.e. $|h_{0,i}|^2 \geq s$.

- All UEs that have successfully decoded the message in the first phase retransmit this message in the second (D2D) phase of $(1-q)T$ channel uses. User j decodes successfully at the end of the second phase if:

$$(1-q)\log_2\left(1+\Big|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\Big|^2\right) \geq q\log_2(1+s),$$

  which is equivalent to:

$$\Big|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\Big|^2 \geq (1+s)^{\frac{q}{1-q}} - 1 \triangleq f_q(s).$$

  where $Z_1(s),\ldots,Z_n(s)$ are binary variables with $Z_j(s)=1$ if j has decoded successfully during the first time slot and 0 otherwise.

In this setting our scheme has two input parameters $s \geq 0$ and $q \in [0,1]$ and we denote by $P_i(s,q)$, $\bar{P}(s,q)$ and $P_+(s,q)$ the ***individual success probability*** of UE i, the *average success probability* of a UE chosen at random and the *joint success probability* for all the UEs, respectively. The performance measures may be written as follows. The average multicast rate is:

$$\begin{aligned} R^m = \quad & \underset{s,q}{\text{maximize}} \quad q\log_2(1+s)\bar{P}(s,q) \\ & \text{subject to} \quad s \geq 0 \text{ and } q \in [0,1]. \end{aligned}$$

which is the expected number of bits received by a UE chosen uniformly at random per channel use, and the multicast outage rate R^o is:

$$\begin{aligned} R^o = \quad & \underset{s,q}{\text{maximize}} \quad q\log_2(1+s) \\ & \text{subject to} \quad P_+(s,q) = 1-\epsilon\,,\ s \geq 0 \text{ and } q \in [0,1]. \end{aligned}$$

It is also noted that this extended model corresponds to our basic model when one sets $q=\frac{1}{2}$.

3.5.1.1 Average Multicast Rate

Individual success probability We first derive the scaling of the average multicasting rate in the regime of a large number of UEs, $n \to \infty$, for any value of $q \in [0,1]$.

Proposition 3.5.1. *For any $i = 1, \dots, n$ and $s \geq 0$ we have:*

$$P_i(s,q) = 1 - (1 - e^{-\frac{s}{\gamma_{0,i}}})\mathbb{E}\left(1 - \exp\left\{-\frac{f_q(s)}{X_i(s)}\right\}\right),$$

where $Z_1(s), \dots, Z_n(s)$ are independent random variables in $\{0,1\}$ with $\mathbb{E}(Z_i(s))$ and $X_i(s)$ given by (3.4) *and* (3.3) *respectively.*

Proof. Consider UE i. Conditionally to the value of $\mathbf{Z}(s)$, UE i does not decode successfully in the first phase if and only if $Z_i(s) = 0$. If $Z_i(s) = 0$, he does not decode successfully in the second phase if and only if $\left|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\right|^2 \geq f_q(s)$. Conditional to $\mathbf{Z}(s)$, random variable $|\sum_{j=1;j\neq i}^{n} h'_{j,i}Z_j(s)|^2$ has exponential distribution with mean $\sum_{j=1;j\neq i}^{n} \gamma_{j,i}Z_j(s) = X_i(s)$, since $\mathbf{Z}(s)$ is independent of $\mathbf{H}'$. Hence

$$\mathbb{P}(i \text{ does not decode}|\mathbf{Z}(s)) = (1 - Z_i(s))(1 - e^{-\frac{f_q(s)}{X_i(s)}}).$$

Taking expectations over $Z_i(s)$ yields the result. □

Asymptotic behavior: We now derive the scaling of the average multicast rate. To do so, we first derive the limiting value of the *individual success probability*, $P_i(\beta \ln n, q)$, for any UE i and any fixed value of $q < 1$. We then argue that the same phase transition phenomenon occurs as in the case of $q = \frac{1}{2}$ considered in Theorem 3.5.1. We conclude that, choosing $s = \beta \ln n$, with $\beta < \beta^\star$ and any fixed $q < 1$ guarantees the following *average success probability* $\bar{P}(s,q) \underset{n\to\infty}{\to} 1$, therefore the average multicast rate scales at least as $(1 - o(1)) \log_2(1 + \beta \ln n)$ when $n \to \infty$. Therefore, optimizing the resource enables to double the achievable rate of our scheme in the regime of a large number of UEs.

Theorem 3.5.1. *(i) For any $\beta > 0$, any $q \in (0,1)$ and $i = 1, \dots, n$, the individual success probability is:*

$$P_i(\beta \ln n, q) \underset{n\to\infty}{\to} \begin{cases} 1 & \text{if } \beta < \beta_{c_i} \\ 0 & \text{otherwise.} \end{cases}$$

(ii) and for any $q \in (0,1)$, the average success probability is:

$$\bar{P}(\beta \ln n, q) \underset{n\to\infty}{\to} \sum_{c=1}^{C} \alpha_c \mathbb{1}\{\beta < \beta_c\},$$

Proof. Consider $0 < q < 1$ fixed, from Proposition 3.5.1:

$$P_i(\beta \ln n, q) = 1 - (1 - n^{-\frac{\beta}{g_{0,c_i}}})\mathbb{E}(1 - e^{-\frac{f_q(\beta \ln n)}{X_i(\beta \ln n)}}).$$

We have $n^{-\frac{\beta}{g_{0,c_i}}} \underset{n\to\infty}{\to} 0$ so that:

$$\lim_{n\to\infty} P_i(\beta \ln n, q) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{f_q(\beta \ln n)}{X_i(\beta \ln n)}}).$$

The $\beta > \beta_{c_i}$ case. Assume that $\beta > \beta_{c_i}$, and we control the expectation of $X_i(\beta \ln n)$. Since $\mathbb{E}(Z_j(s)) = e^{-\frac{s}{g_{0,c_j}}}$:

$$\begin{aligned}
\mathbb{E}\Big(\frac{X_i(\beta \ln n)}{f_q(\beta \ln n)}\Big) &= \sum_{j=1;j\neq i}^{n} \frac{\gamma_{j,i} n^{-\frac{\beta}{\gamma_{0,j}}}}{f_q(\beta \ln n)} \\
&\leq \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}{f_q(\beta \ln n)} \\
&\leq \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{\beta_{c_i}}}}{f_q(\beta \ln n)} \underset{n\to\infty}{\to} 0
\end{aligned}$$

since for all c either $g_{c,c_i} = 0$ or $g_{0,c} \leq \beta_{c_i}$, and $\beta > \beta_{c_i}$ and $f_q(\beta \ln n) \underset{n\to\infty}{\sim} (\beta \ln n)^{\frac{q}{1-q}}$. Therefore, $\frac{X_i(\beta \ln n)}{f_q(\beta \ln n)}$ converges to 0 in L^1 so that it converges to 0 in distribution as well. Since $x \mapsto e^{-\frac{1}{x}}$ is both continuous and bounded we get:

$$\lim_{n\to\infty} P_i(\beta \ln n, q) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{f_q(\beta \ln n)}{X_i(\beta \ln n)}}) = 0.$$

The $\beta < \beta_{c_i}$ case. Now consider $\beta < \beta_{c_i}$.

It is noted that (3.5) still holds in this context (see Theorem 3.5.1), and since $x \mapsto e^{-\frac{1}{x}}$ is increasing:

$$\mathbb{E}(e^{-\frac{f_q(\beta \ln n)}{X_i(\beta \ln n)}}) \geq \mathbb{P}\left(X_i(\beta \ln n) \geq \frac{\mathbb{E}(X_i(\beta \ln n))}{2}\right) e^{-\frac{2 f_q(\beta \ln n)}{\mathbb{E}(X_i(\beta \ln n))}}.$$

Consider $\hat{c}$ such that $g_{\hat{c},c_i} > 0$ and $g_{0,\hat{c}} = \beta_{c_i}$. Then:

$$\begin{aligned}
\mathbb{E}\left(\frac{X_i(\beta \ln n)}{f_q(\beta \ln n)}\right) &= -\frac{\gamma_{i,i} e^{-\frac{s}{\gamma_{0,i}}}}{f_q(\beta \ln n)} + \sum_{c=1}^{C} \frac{\alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}{f_q(\beta \ln n)} \\
&\geq -\frac{\gamma_{i,i} e^{-\frac{s}{\gamma_{0,i}}}}{(\beta \ln n)^{\frac{q}{1-q}} - 1} + \frac{\alpha_{\hat{c}} g_{\hat{c},c_i} n^{1-\frac{\beta}{\beta_{c_i}}}}{(\beta \ln n)^{\frac{q}{1-q}} - 1} \underset{n\to\infty}{\to} \infty
\end{aligned}$$

Replacing in (3.5) we deduce

$$\mathbb{P}\left(X_i(\beta\ln n) \geq \frac{\mathbb{E}(X_i(\beta\ln n))}{2}\right) \underset{n\to\infty}{\to} 1,$$

and $e^{-\frac{2f_q(\beta\ln n)}{\mathbb{E}(X_i(\beta\ln n))}} \underset{n\to\infty}{\to} 1$ so that:

$$\lim_{n\to\infty} P_i(\beta\ln n, q) = \lim_{n\to\infty} \mathbb{E}(e^{-\frac{f_q(\beta\ln n)}{X_i(\beta\ln n)}}) = 1.$$

which completes the proof of statement (i). Statement (ii) follows from the fact that $\bar{P}(s) = \frac{1}{n}\sum_{i=1}^{n} P_i(s)$. □

Corollary 3.5.2. *For any $\beta < \beta^\star$ we have:*

$$R^m \geq (1 - o(1))\log_2(1 + \beta\ln n)\,,\ n \to \infty.$$

Non asymptotic behavior As done for the basic case of $q = \frac{1}{2}$, we may derive an asymptotic approximation of the *individual success probability*. The proof is similar to that of Theorem 3.3.3 and is omitted.

Theorem 3.5.3. *Consider $\beta < \beta_c$, and any $q \in (0,1)$ then:*

$$1 - P_i\left(\beta\ln n, q\right) \underset{n\to\infty}{\sim} 1 - \exp\left\{\frac{-(\beta\ln n)^{\frac{q}{1-q}}}{\sum_{c=1}^{C}\alpha_c g_{c,c_i} n^{1-\frac{\beta}{g_{0,c}}}}\right\}.$$

3.5.1.2 Multicast Outage Rate

Joint success probability

Proposition 3.5.2. *For any $s \geq 0$ we have:*

$$P_+(s,q) = \mathbb{E}\left[\exp\left\{-f_q(s)\sum_{i=1}^{n}\frac{1 - Z_i(s)}{X_i(s)}\right\}\right],$$

where $Z_1(s),\ldots,Z_n(s)$ are independent random variables in $\{0,1\}$ and $\mathbb{E}(Z_i(s)) = e^{-\frac{s}{\gamma_{0,i}}}$ for $i = 1,\ldots,n$ and $X_i(s) = \sum_{j=1;j\neq i}^{n} Z_j(s)\gamma_{j,i}$.

Proof. We use the same notation as in the proof of Proposition 3.3.1. User i successfully decodes if and only if either $Z_i(s) = 1$ or $|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}|^2 \geq f_q(s)$. Hence i decodes if and only if $\left|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\right|^2 \geq f_q(s)(1 - Z_i(s))$. Now:

$$\begin{aligned}\mathbb{P}(\text{all decode}|\mathbf{Z}(s)) &= \mathbb{P}\Big(\Big|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\Big|^2 \geq f_q(s)(1 - Z_i(s))\,,\ \forall i|\mathbf{Z}(s)\Big)\\ &= \prod_{i=1}^{n}\mathbb{P}\Big(\Big|\sum_{j=1;j\neq i}^{n} Z_j(s)h'_{j,i}\Big|^2 \geq f_q(s)(1 - Z_i(s))|\mathbf{Z}(s)\Big)\\ &= \exp\left\{-f_q(s)\sum_{i=1}^{n}\frac{1 - Z_i(s)}{X_i(s)}\right\}.\end{aligned}$$

since conditional to $\mathbf{Z}(s)$, $|\sum_{j=1}^{n} Z_j(s)h'_{j,i}|^2$ has exponential distribution with mean $X_i(s)$, and $\mathbf{H}'$ has independent entries. Averaging over $\mathbf{Z}(s)$ yields the result. □

Asymptotic behavior

Theorem 3.5.4. *For any $s \geq 0$ we have:*

$$P_+(s,q) \underset{n\to\infty}{\to} \exp\left\{-\sum_{c=1}^{C}\frac{f_q(s)\alpha_c(1 - e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C}\alpha_{c'}g_{c',c}e^{-\frac{s}{g_{0,c'}}}}\right\}.$$

Corollary 3.5.5. *When $n \to \infty$, the multicast outage rate converges $Q(s)\log_2(1+s)$ where $Q(s)$ is the unique solution to:*

$$((1+s)^{\frac{Q(s)}{1-Q(s)}} - 1)\sum_{c=1}^{C}\frac{\alpha_c(1 - e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C}\alpha_{c'}g_{c',c}e^{-\frac{s}{g_{0,c'}}}} = \ln\Big(\frac{1}{1-\epsilon}\Big).$$

$$Q(s) = 1 - \left(\frac{\ln\left(1 - \ln(1-\epsilon)F(s)^{-1}\right)}{\ln(1+s)} + 1\right)^{-1}$$

where $F(s) = \sum_{c=1}^{C}\frac{\alpha_c(1-e^{-\frac{s}{g_{0,c}}})}{\sum_{c'=1}^{C}\alpha_{c'}g_{c',c}e^{-\frac{s}{g_{0,c'}}}}$

3.5.2 HARQ

We now consider the case where UEs decode by combining the signals received in two phases using some form of HARQ. Indeed, in the baseline scheme studied in the previous sections, if a UE does not decode successfully in the first phase it simply discards the received signal from the first phase. Although the proposed scheme enhanced by HARQ yields better performance by decreasing further the decoding error probability, the following analysis reveals that it does not improve the scaling law in the regime of a large number of UEs. For simplicity we assume that there is no resource optimization between the two successive phases, so that $q = \frac{1}{2}$. Our results can be extended to the cases where $q \neq \frac{1}{2}$ in a straightforward manner.

Using the same notation, UE i decodes if and only if the following conditions are met:

$$\log_2\left(1 + |h_{0,i}|^2 + \Big|\sum_{j=1;j\neq i}^{n} Z_i(s)h'_{j,i}\Big|^2\right) \geq \log_2(1+s),$$

for CC and

$$\log_2(1 + |h_{0,i}|^2) + \log_2\left(1 + \Big|\sum_{j=1;j\neq i}^{n} Z_i(s)h'_{j,i}\Big|^2\right) \geq \log_2(1+s).$$

for IR.

For the average multicast rate, let us consider two random variable ξ_1 and ξ_2 both following an exponential law of parameter 1. When using the HARQ scheme with CC, the condition for decoding is the following:

$$\log_2(1 + \gamma_{0,i}\xi_1 + X_i(s)\xi_2) \geq \log_2(1+s)$$

or equivalently

$$\frac{\gamma_{0,i}\xi_1}{s} + \frac{X_i(s)\xi_2}{s} \geq 1 \tag{3.6}$$

Hence the *individual success probability* of UE i when using HARQ with CC is

$$P_i^{CC}(s) = \mathbb{P}\left(\frac{\gamma_{0,i}\xi_1}{s} + \frac{X_i(s)\xi_2}{s} \geq 1\right)$$

When using the HARQ scheme with IR, the condition for successful decoding is the following:

$$\log_2(1 + \gamma_{0,i}\xi_1) + \log_2(1 + X_i(s)\xi_2) \geq \log_2(1+s)$$

or equivalently

$$\frac{\log_2(1+\gamma_{0,i}\xi_1)}{\log_2(1+s)} + \frac{\log_2(1+X_i(s)\xi_2)}{\log_2(1+s)} \geq 1 \tag{3.7}$$

Hence the *individual success probability* of UE i when using HARQ with CC is

$$P_i^{IR}(s) = \mathbb{P}\left(\frac{\log_2(1+\gamma_{0,i}\xi_1)}{\log_2(1+s)} + \frac{\log_2(1+X_i(s)\xi_2)}{\log_2(1+s)} \geq 1\right)$$

Let us consider $s = \beta \ln n$, in the regime of $n \to \infty$,

$$\frac{\gamma_{0,i}}{\beta \ln n} \underset{n\to\infty}{\to} 0$$

implying that the first term in the left hand side (l.h.s.) of (3.6) and (3.7) disappears so that the benefit of HARQ becomes negligible. Hence for HARQ CC, the *individual success probability* is:

$$P_i^{CC}(\beta \ln n) \underset{n\to\infty}{\to} \mathbb{P}\left(X_i(\beta \ln n)\xi_2 \geq \beta \ln n\right) = P_i(\beta \ln n)$$

And for HARQ IR, *individual success probability* is:

$$P_i^{IR}(\beta \ln n) \underset{n\to\infty}{\to} \mathbb{P}\left(X_i(\beta \ln n)\xi_2 \geq \beta \ln n\right) = P_i(\beta \ln n)$$

Corollary 3.5.6. *Theorem 3.3.1 can be applied to HARQ CC or IR, and*

$$P_i^{HARQ}\left(\beta \ln n\right) \underset{n\to\infty}{\to} \begin{cases} 1 & \textit{if } \beta < \beta_{c_i} \\ 0 & \textit{otherwise.} \end{cases}$$

3.6 Numerical Experiments

In this section we present numerical experiments and show that our theoretical analysis provides accurate predictions of the system's behavior.

We consider two scenarios:

(1) A single class $C = 1$ scenario, with $\alpha = (1)$, $g_{0,1} = 36.5$ and $g_{1,1} = 28.8$, representing the homogeneous case where the UEs are close to the BS and to each other.

(2) A multi-class scenario represented by a cell of radius 250 m and n UEs drawn uniformly at random. $g_{i,j}$ is given by the ratio between the received signal power and the noise power, where the received power by UE i is represented by $\rho_i - (128,1 + 37.6\log_{10}(d_{i,j}))$ dBm where the transmitted power is $\rho_i = 46$ dBm if $i = 0$ and $\rho_i = 23$ dBm otherwise and $d_{i,j}$ is the distance between i and j, expressed

in kilometers. The noise at the UE side is given by $10\log_{10}(W10^{-17,4})$ dBm where the frequency band W is taken as 20 MHz with a white noise power density of -174 dBm/Hz, and a noise figure of 5 dB if $i = 0$ and 9 dB otherwise. This scenario represents the case where UEs arrive uniformly in a cell area, which seems like an acceptable model for a real system in high load (i.e. when there are many active UEs).

For both scenarios $\epsilon = 10^{-2}$. All the chosen parameters are given by the 3GPP standards [34].

3.6.1 Main results

The results presented in this section refer to both Section 3.3 and 3.4. For each figure, three curves are presented: "baseline" is the performance of the baseline scheme, "simulation" is the exact performance of the proposed scheme obtained by simulation, and "approx" is the analytical approximation of the proposed scheme's performance. For the average multicast rate the approximation is given by Theorem 3.3.3 and for the multicast outage rate, the approximation is given by Theorem 3.4.1.

For the average multicast rate, Fig. 3.2 presents the average multicast rate R^m as a function of the number of UEs n (when s is chosen optimally), while Fig. 3.3 presents the average multicast rate as a function of s, for various values of n. For multicast outage rate, Fig. 3.4 presents the multicast outage rate as a function of the number of UEs n, while Fig. 3.5 presents the error probability $1 - P_+(s)$ as a function of s for system size $n = 100$.

These numerical experiments point out three facts:

- The proposed scheme is vastly superior to the baseline scheme as soon as there are more than a few UEs, as predicted by our asymptotic analysis.

- The proposed approximations/asymptotic expressions derived above predict the performance of the proposed scheme very accurately, even for systems of modest size, say $n \geq 50$ for the average multicast rate and $n \geq 10$ for the multicast outage rate.

- As a consequence, setting the parameter s can be done in a simple and tractable manner to obtain good practical performance for systems of modest size.

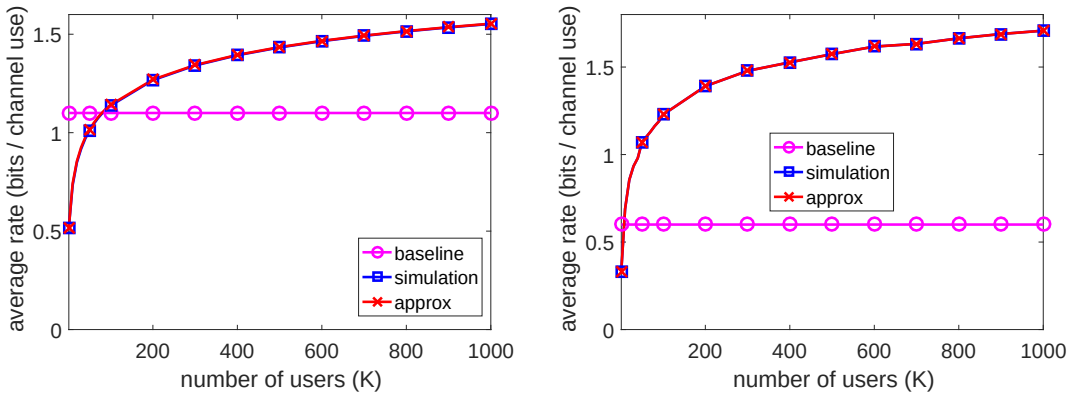


Figure 3.2: Average multicast rate R^m versus number of UEs n.

Scenario (1) (left), Scenario (2) (right).

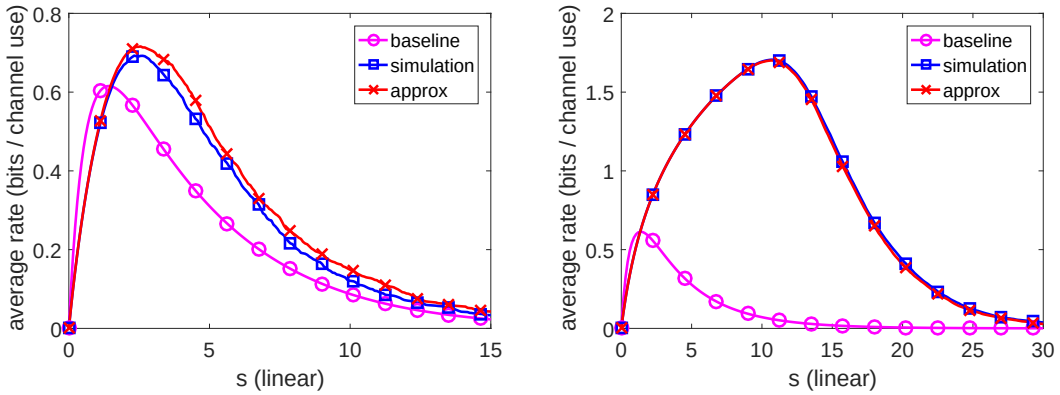


Figure 3.3: Average multicast, effective rate $\frac{1}{2}\log_2(1+s)\bar{P}(s)$ vs s.

Scenario (2). $n = 10$ (left), $n = 10^3$ (right).

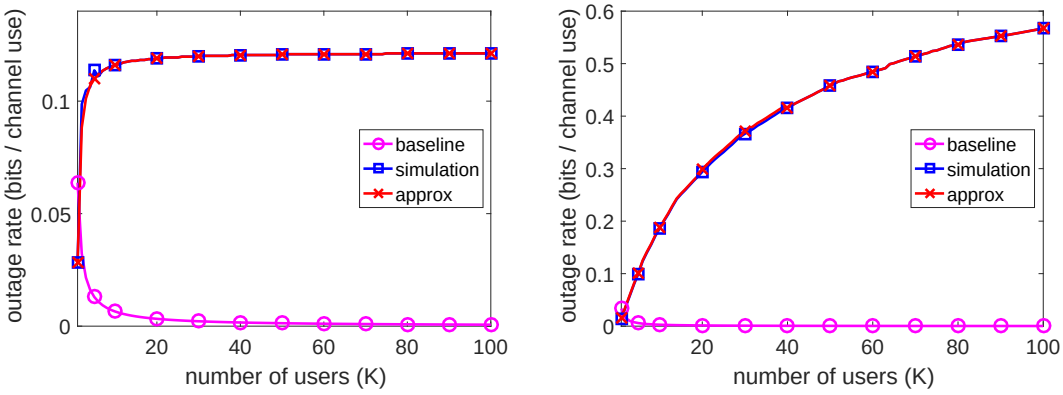


Figure 3.4: Multicast outage rate R^o versus number of UEs n.

Scenario (1) (left), Scenario (2) (right).

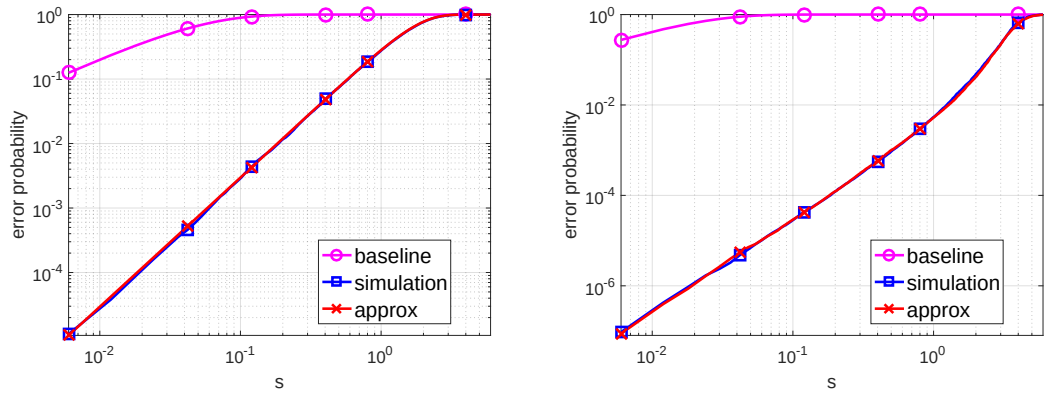


Figure 3.5: Multicast outage rate, probability of error $1 - P_+(s)$ vs s for $n = 10^2$ UEs. Scenario (1) (left), Scenario (2) (right).

3.6.2 Resource allocation

The results presented in this section refer to Section 3.5.1 where one is allowed to optimize resource utilization across the two phases of our main scheme. The previous "simulation" curve from Fig. 3.2 to 3.5 is now called "$q = 0.5$" as it corresponds to the case where the time slots of the two phases are of same duration. In this sense, three curves are now compared: "baseline" is the performance of the baseline scheme, "$q = 0.5$" obtained by simulation, and "optimal q" which represents the exact average multicast rate performance of the proposed scheme when optimizing the resource value q. For the optimization resource, the average multicast rate is deduced from Proposition 3.5.1. For the average multicast rate, Fig. 3.6 presents the average multicast rate R^m and the optimal resource q as a function of the number of UEs n (when s is chosen optimally), while Fig. 3.7 presents the average multicast rate as a function of s, for various values of n. For multicast outage rate, Fig. 3.8 presents the multicast outage rate as a function of the number of UEs n and the optimal resource q as a function of s. These numerical experiments point out four facts:

- The proposed resource allocation scheme provides higher average multicast rate compared to the scheme with equal repartition of the resource, $q = 0.5$.
- The optimal resource associated to the average multicast scheme is increasing with the number of UEs as predicted by Corollary 3.5.2.
- There is a trade-off between resource optimization and SNR selection when the number of UEs is fixed. Indeed, we select a smaller SNR when using the resource allocation scheme while achieving a higher average multicast rate.

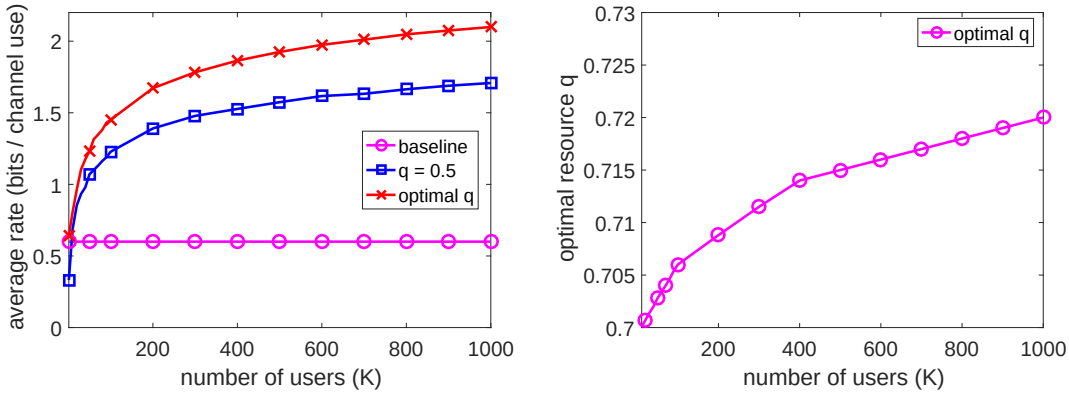


Figure 3.6: Resource optimization for average multicast rate R^m versus number of UEs. Scenario (2). Multicast rate R^m (left), Optimal resource allocation (right).

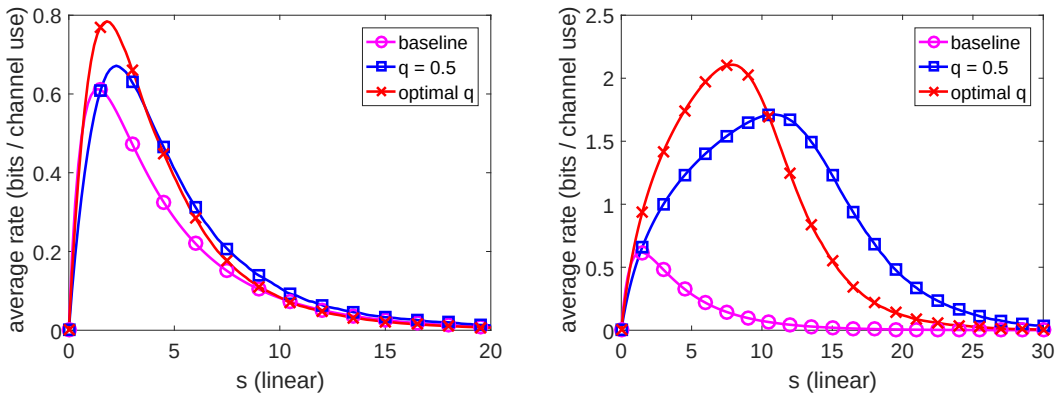


Figure 3.7: Average multicast effective rate vs s with resource allocation. Scenario (2). $n = 10$ (left), $n = 10^3$ (right)

- For the multicast outage rate, the use of resource optimization provides a quickest convergence to the asymptotic regime compared to the basic scheme ($q = 0.5$).

3.6.3 HARQ

The results presented in this section refer to Section 3.5.2. For each figure, three curves are presented: "simulation" is the exact performance of the main proposed scheme ($q = 0.5$) obtained by simulation, "HARQ CC" and "HARQ IR" are the exact performance of our scheme also obtained by simulation when adding HARQ CC and HARQ IR respectively. For the average multicast rate, the *individual*

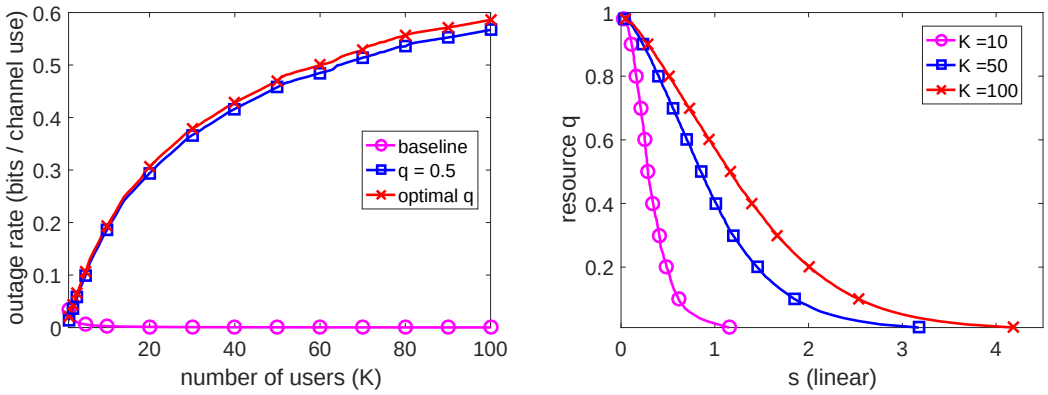


Figure 3.8: Resource optimization for multicast outage rate.

Scenario (2). Multicast outage rate R^o versus number of UEs n (left). Optimal resource for different number of UEs (right).

success probability for UE i is derived for both HARQ schemes in Section 3.5.2 and is given by Corollary 3.5.6. Fig. 3.9 presents the average multicast rate R^m as a function of the number of UEs n (when s is chosen optimally), while Fig. 3.10 presents the average multicast rate as a function of s, for various values of n. These numerical experiments point out one more fact: both HARQ schemes provide better performance than without HARQ.

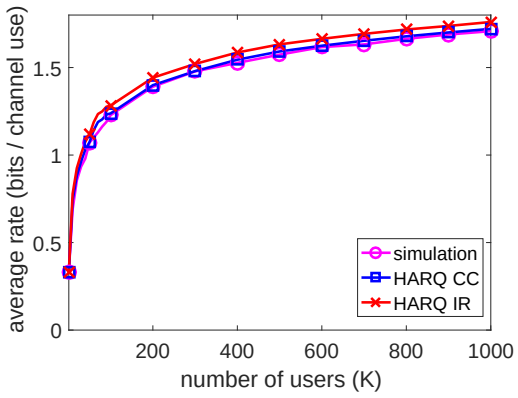


Figure 3.9: Average multicast rate R^m versus number of UEs n with HARQ.

Scenario (2).

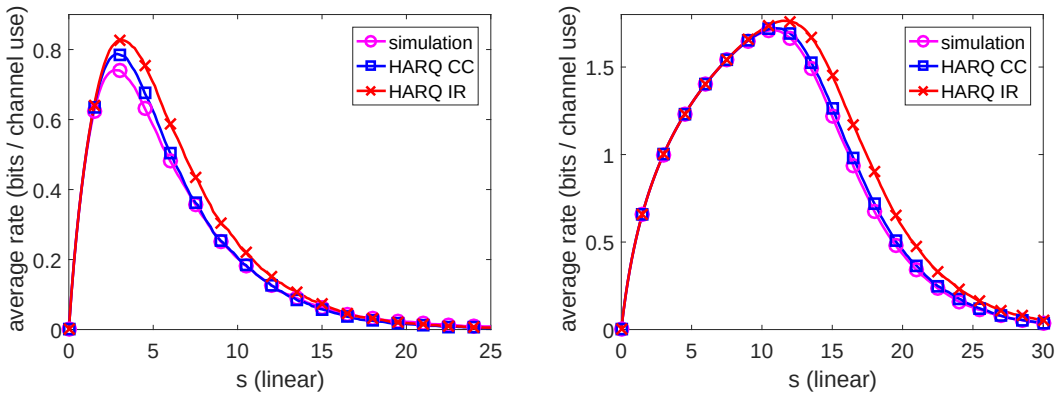


Figure 3.10: Average multicast effective rate vs s with HARQ.

Scenario (2). $n = 10$ (left), $n = 10^3$ (right).

3.7 Conclusion

In this chapter, we have considered the multicast wireless scenario aided by D2D communication. This scenario is important and not limited to the introduction of multimedia mobile video streaming or PS in the future 5G wireless network. Enhancement in terms of performance has been proposed for scenarios like multimedia mobile video streaming (average multicast rate) and PS (multicast outage rate). In particular, we have studied a new scheme for multicasting a message to UEs with receiver-to-receiver communication, without channel state information at the transmitter. We have shown that, in the large system limit, the performance of this scheme can be completely characterized. For finite systems, we also have provided tractable approximations for performance measures which are very accurate even for systems of modest sizes. We also provide optimization of the resource utilization in a first extension and consider the use of HARQ in a second extension for our general scheme.

3.8 Appendix

We recall below the list of parameters used along the chapter:

Table 3.1: List of parameters in Chapter 3

n	Number of UEs
$y_i[t]$	received signal at UE i
$x_i[t]$	transmitted signal by UE i (BS is indexed by 0)
$N_i[t] \sim \mathcal{N}(0,1)$	AWGN at UE i
$\mathbf{h} = (h_{0,i})_{i=1,\dots,n}$	vector of channel coefficients during the first time slot
$\mathbf{H}' = (h'_{i,j})_{i,j=1,\dots,n}$	matrix of channel coefficients during the second time slot
$\mathbf{\Gamma} = (\gamma_{i,j})_{i,j=0,\dots,n}$	matrix of channel statistics gain between UEs
s	SNR threshold parameter chosen depending on $\mathbf{\Gamma}$
$\log_2(1+s)$	transmitted rate from the BS
$\lvert h_{0,i}\rvert^2 \sim \exp(\frac{1}{\gamma_{0,i}})$	received SNR at UE i during the first time slot
$Z_i(s) \overset{\Delta}{=} \mathbb{1}_{\{\lvert h_{0,i}\rvert^2 > s\}}$	binary outcome of UE's i success in the first time slot
$X_i(s) \overset{\Delta}{=} \sum_{j=1}^n Z_j(s)\gamma_{j,i}$	sum channel gain of the UEs transmitting in the second time slot
$\lvert \sum_{j=1}^n Z_j(s) h'_{j,i}\rvert^2 \sim \exp(\frac{1}{X_i(s)})$	received SNR at UE i during the second time slot
c_i	class of UE i
C	number of classes
α_c	UEs' proportion of class c
$(g_{c_i,c_j})_{c_i,c_j=0,\dots,C}$	mean channel gain between classes
G	max mean channel gain
β_c	largest value of $g_{0,c'}$ from BS to class c' which communicates with class c in the second phase
$\beta^\star$	minimum value of β_c
q	time coefficient of the first time slot
P_i	individual success probability
$\bar{P}$	average success probability
P_+	joint success probability
$\varepsilon = 10^{-2}$	error of P_+
R^m	average multicast rate
R^o	multicast outage rate

Chapter 4

Device-to-Device aided Ultra Reliable Low Latency Communication

In this chapter we consider the communication under URLLC constraint, i.e. latency and reliability. Aside from being a general and simple model to study the impact of delayed feedback on HARQ schemes, our setting is motivated by several features present in current wireless networking systems and/or their near future evolution), such as URLLC. These systems require low latency, so that the time slot length must be short, and feedback delay cannot be assumed to be negligible with respect to the slot length. These systems must be highly reliable (fraction of packet lost should be less than 1 out of 10^5), so that retransmissions cannot be avoided due to the noisy nature of the wireless medium. Also, one notices that if there are no deadline constraints, the optimal policy is to wait for feedback after each transmission, so the problem becomes trivial. Deadlines appear naturally in several applications such as guiding systems for autonomous cars. The objective is also natural: minimizing the number of retransmissions corresponds to minimizing the average power consumption subject to a reliability constraint. Further, if the wireless medium is shared with other systems (in cognitive radio fashion), minimizing the number of transmissions also leads to minimizing the amount of interference created to other systems. In the case of multiple UE transmitters, we consider distributed protocols since implementing a central scheduler is expensive, and does not scale to realistic scenarios where the number of communicating objects is envisioned to be very large. We consider fixed delay for simplicity and also due to the fact that the feedback jitter should be small, as delay is caused mostly by physical transmission of signals and processing at the UE receiver.

In this chapter, we argue that both approaches (waiting for a NACK before retransmitting versus blind retransmissions) are not incompatible and an optimal policy has to consider both possibilities. In particular, even if blind retransmissions of replicas are performed under a contention-based scheme with no individual

reservation, the UE receiver may successfully decode the packet, discover its origin and send an ACK to the UE transmitter. This will result in stopping the transmissions; otherwise, retransmissions may continue until the latency budget expires. Deriving the optimal policy in this case is not trivial, as this corresponds to a decision process with delayed feedback. The original contributions of this paper are the following:

- In the case of one-to-one transmission, we analyze the optimal HARQ schemes for URLLC services when the feedback arrives with a delay that is larger than the transmission interval (as it is usually the case). We derive the optimal policies and show that they can be computed in reasonable time.
- We propose a novel semi-greedy policy that has the advantage of being simple, of relying only on the knowledge of the service requirements and the channel feedback delay (and not the error rate), and of being optimal in many practical cases that we identify. We also propose an extension to the multicast case of a single transmitter and several receivers with the same type of wireless errors.
- We extend our results to the system level where there are multiple UE transmitters competing for the channel access and requiring URLLC service. We derive retransmission policies that meet the URLLC targets with the lowest number of retransmissions.

The remainder of this chapter is organized as follows. In Section 4.1, we describe two variants of the system model, one corresponding to the link level, where a single UE transmitter transmits packets to a single UE receiver on a noisy channel, and another corresponding to the system level, where a multitude of UE transmitters contend on a common channel for reaching a single UE receiver. Low complexity yet robust retransmission policies for the single UE transmitter and multiple UE transmitters cases are derived in Sections 4.2 and 4.3, respectively. Section 4.4 concludes the chapter.

4.1 System Model

We consider two variants for the problem of HARQ with delayed feedback where UE transmitters wish to convey packets to a single UE receiver subject to a hard deadline. The UE receiver may fail to decode the received packet correctly, so that a simple ACK/NACK feedback is used, and the packet can be retransmitted to ensure high reliability. The ACK/NACK feedback is received with some fixed delay, so that if one transmits a packet once, one may retransmit it another time without knowing whether or not the first transmission was successful. When the

feedback delay is known, there is a clear trade-off between (a) retransmitting the packet in bursts without waiting for the feedback which ensures high reliability but many spurious retransmissions and (b) waiting for the feedback of each transmission which yields no spurious retransmissions but may fail to transmit the packet enough times for it to be successfully decoded before the deadline.

We both consider (i) the case of a single UE transmitter, where the UE receiver may fail to decode the packet due to an external source of noise e.g. thermal noise, or another wireless system operating on the same radio resources, and (ii) the case where multiple UEs transmit using some distributed access scheme, and the UE receiver fails to decode the packet if and only if there is a collision.

4.1.1 Single UE transmitter model

Link model. We consider a single UE transmitter ($n = 1$) whom attempts to transmit one packet to a single UE receiver ($N = 1$). Time is slotted and there are a finite number D of slots after the packet is generated before the delay constraint is violated. At each slot $t \in \{1, \dots, D\}$ the UE may choose to transmit or not where $X_t = 1$ if it transmits and $X_t = 0$ otherwise. If the UE transmits at slot t, the UE receiver successfully decodes the packet with probability p. Let $Y_t = 1$ if the UE receiver succeeds and $Y_t = 0$ if it fails to decode. Then the packet is successfully delivered at slot t if and only if $X_t Y_t = 1$. If the packet is not delivered after D slots (the deadline) so that $\sum_{t=1}^{D} X_t Y_t = 0$, it is considered lost. We assume that the success probability of a transmission is $p \in [0, 1]$, and that successes/failures are i.i.d. namely $Y_1, \dots, Y_D$ are i.i.d Bernoulli variables with expectation $\mathbb{E}(Y_t) = p$.

Feedback model. Whenever the UE transmits, it received feedback indicating whether or not the packet has been successfully decoded after a fixed delay of Δ time slots[1]. Namely, if it transmits at slot t it receives $X_t Y_t$ at time $t' = t + \Delta$. Denote by Z_t the feedback received at time t, then we have

$$Z_t = \begin{cases} X_{t-\Delta} Y_{t-\Delta} & \text{if } t \geq \Delta + 1, \\ 0 & \text{if } t \leq \Delta. \end{cases}$$

Policies and objective. The UE transmitter must sequentially choose whether or not to transmit at each slot based on the received feedback, so that X_t, the decision to transmit at t must depend only on the feedback available at that time $Z_1, \dots, Z_t$. Formally, under policy π we have:

$$X_t^\pi = f_t^\pi(Z_1, \dots, Z_t),$$

[1]The explicit generation of a NACK is not mandatory as the absence of an ACK after a time Δ can be interpreted as a NACK. This is particularly the case when the UE receiver does not even detect the presence of a packet to decode.

where X_t^π is the decision at time t under policy π, and $f_t^\pi : \{0,1\}^t \to \{0,1\}$ is a function. We define Π the set of all possible policies. Our objective, denoted by problem (P_1), is to minimize the number of transmissions, subject to a reliability constraint, where the probability of not successfully decoding the packet is equal to a known threshold:

$$\begin{aligned} &\underset{\pi\in\Pi}{\text{minimize}} \sum_{t=1}^{D} \mathbb{E}(X_t^\pi) && (P_1)\\ &\text{subject to } \mathbb{P}\left(\sum_{t=1}^{D} X_t^\pi Y_t = 0\right) = (1-p)^m \end{aligned}$$

where $m \in \{0,\dots,D\}$ depends on the reliability threshold ϵ: $m = \lceil \frac{\ln(\epsilon)}{\ln(1-p)} \rceil$. Two remarks are in order. First the reliability, that cannot exceed $(1-p)^D$, is chosen among the possible values $(1-p)^m$, $m \in \{0,\dots,D\}$. All policies that have the same number of transmitted replicas, m, have the same reliability of a blind policy which always transmits m times and completely disregards feedback, since each transmission fails with probability $1-p$ and failures are independent across time slots. Of course, this blind policy is usually far from optimal since it disregards feedback and potentially leads to a large number of spurious transmissions. Secondly problem (P_1) is a stochastic control problem with a chance constraint, and this problem can be complex to solve in the general case. We denote by V the optimal value of (P_1), and $\pi^\star \in \Pi$ an optimal policy of (P_1). We study how to solve problem (P_1) in details in Section 4.2.

4.1.2 Multiple UE transmitters model

Link model. In our second model, there are $n \in \{1,\dots,D\}$ UE transmitters whom share a common wireless channel. Each UE wants to convey a single packet to a common UE receiver. As previously, there are D time slots after the generation of a packet, and at any time slot $t \in \{1,\dots,D\}$ we write $X_{t,i} = 1$ if UE $i \in \{1,\dots,n\}$ transmits and $X_{t,i} = 0$ otherwise. In this model the UE receiver successfully decodes if and only if there are no collisions. Namely, the UE receiver successfully decodes at time t if and only if exactly one UE has transmitted, so that $\sum_{i=1}^{n} X_{t,i} = 1$. Since UE i has only one packet to transmit to the UE receiver, we consider that i is successful if and only if at least one of its transmissions has been decoded without collision, i.e. $\sum_{t=1}^{D} \sum_{j\neq i} X_{t,i}(1 - X_{t,j}) \geq 1$.

Feedback model. As before, we consider delayed feedback indicating whether or not a collision has occurred. Transmitter i observes a feedback of 1 (denoting a collision) at time $t+\Delta$ if and only if i has transmitted at time t, and there exists another UE $j \neq i$ whom also transmitted at t, and the feedback is 0 otherwise.

Denote by $Z_{t,i}$ the feedback received at time t by UE transmitter i then we have

$$Z_{t,i} = \begin{cases} X_{t-\Delta,i} \sum_{j \neq i} X_{t-\Delta,j} & \text{if } t \geq \Delta + 1, \\ 0 & \text{if } t \leq \Delta. \end{cases}$$

Policies and objective. Each UE transmitter must sequentially choose whether or not to transmit at each slot based on the received feedback. We impose two constraints on the policies that UE transmitters may implement: first the policy should be symmetrical so that if two UE transmitters observe the same feedback, they should adopt the same behavior, and second the policy should be distributed so that UE transmitters may not exchange information and base their decisions solely on the observed feedback. We investigate probabilistic decisions, so that a UE may choose whether or not to transmit based on the observed feedback. Formally, under policy π, UE i transmits with probability given by:

$$\mathbb{E}(X_{t,i}^{\pi} | Z_{1,i}, \ldots, Z_{t,i}) = f_t^{\pi}(Z_{1,i}, \ldots, Z_{t,i}).$$

where $f_t^{\pi} : \{0,1\}^t \to \{0,1\}$ is a function which maps the observed feedback at time t into the distribution of $X_{t,i}^{\pi}$. Denote by Π the set of admissible policies. The objective in this model is to maximize the proportion of UE transmitters whom successfully deliver their packet, which corresponds to:

$$\underset{\pi \in \Pi}{\text{maximize}} \ \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}\left(\sum_{t=1}^{D} \sum_{j \neq i} X_{t,i}^{\pi}(1 - X_{t,j}) \geq 1 \right) \qquad (P_2)$$

We propose several distributed policies for this problem and study their performance in section 4.3.

4.2 Single Transmitter

Throughout this section we consider the model with a single UE transmitter described in Subsection 4.1.1 and we study the optimal policy of problem (P_1).

4.2.1 Reduction to an optimization problem

As mentioned earlier, problem (P_1) is a stochastic control problem with a chance constraint, which can be cumbersome to solve. We reduce P_1 to a much simpler *deterministic* optimization problem P_1', which is easier to solve and takes randomness out of the original problem.

Proposition 4.2.1. *Consider the optimization problem:*

$$\underset{\mathbf{x}\in\{0,1\}^D}{\text{minimize}}\ g(\mathbf{x}) \equiv \sum_{t=1}^{D} x_t(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}} \qquad (P_1')$$

$$\text{subject to } \sum_{t=1}^{D} x_t = m.$$

Define $\mathbf{x}^\star$ *an optimal solution of* (P_1') *and define the policy* π *such that*

$$X_t^\pi = x_t^\star \prod_{t'=1}^{t}(1-Z_{t'}).$$

Then π *is optimal for the original problem* (P_1).

Proof. The reduction is based on the fact that all optimal policies verify the following property. These policies transmit at time t if and only if the feedback received up to time t, i.e. $Z_1,\dots,Z_t$ are all 0. Otherwise, if there exists $t' \le t$ such that $Z_{t'} = 1$ the UE transmitter knows that the UE receiver has already decoded successfully before t, and transmitting at time t is unnecessary. Define $\bar{\Pi}$ the set of such policies. Consider $\pi \in \bar{\Pi}$ such a policy. We must have:

$$\begin{aligned} X_t^\pi &= f_t^\pi(Z_1,\dots,Z_t) = f_t^\pi(0,\dots,0)\prod_{t'=1}^{t}(1-Z_{t'}) \\ &= x_t^\pi \prod_{t'=1}^{t}(1-Z_{t'}), \end{aligned}$$

since $f_t^\pi(Z_1,\dots,Z_t) = 0$ unless $Z_1,\dots,Z_t$ are all 0, with $x_t^\pi = f_t^\pi(0,...,0)$. The expected number of transmissions under π is:

$$\begin{aligned} \sum_{t=1}^{D}\mathbb{E}(X_t^\pi) &= \sum_{t=1}^{D} x_t^\pi \mathbb{E}(\prod_{t'=1}^{t}(1-Z_{t'})) \\ &= \sum_{t=1}^{D} x_t(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}} \end{aligned}$$

since π transmits at t if and only if $x_t^\pi = 1$ and all transmission at times $1,\dots,t-\Delta$ have failed. The probability of this event is $(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}}$ since there are $\sum_{t'=1}^{t-\Delta} x_{t'}$ such transmission attempts, and failures are i.i.d. with probability $1-p$. Also, the probability that the packet is not decoded after the deadline D is:

$$\mathbb{P}\left(\sum_{t=1}^{D} X_t^\pi Y_t = 0\right) = (1-p)^{\sum_{t=1}^{D} x_t},$$

since π transmits $\sum_{t=1}^{D} x_t^\pi$ times. Since the l.h.s. must be equal to $(1-p)^m$, by definition of (P_1) this imposes the constraint $\sum_{t=1}^{D} x_t^\pi = m$. The mapping between $\mathbf{x}$ and $\pi \in \bar{\Pi}$ is one-way, solving (P_1) or (P_1') is equivalent which yields the result. □

4.2.2 Optimal policy

We now turn our attention to algorithms in order to solve the reduced problem (P_1') defined in the previous subsection. The number of policies in (P_1) is 2^{D+1}, is much larger than the number of solutions of (P_1') which is simply $\binom{D}{m}$. Still, in the worse case exhaustive search requires $\binom{D}{D/2}$ operations which is exponential in the problem dimension D. We now show that (P_1') can be solved in time which is linear in m and D for a fixed value of the delay Δ. The key idea is that (P_1') can further be reduced to a Markov Decision Process (MDP) with deterministic transitions, a time horizon of D and a number of state-action pairs (at each time) smaller than $m2^\Delta$. For complete explanation of MDP, readers are invited to refer to [110].

Theorem 4.2.1. *The optimal solution of the reduced problem (P_1') may be computed in time $\mathcal{O}(Dm2^\Delta)$ and memory $\mathcal{O}(m2^\Delta)$ using dynamic programming, for any value of p.*

Proof. We show that any solution $\mathbf{x} \in \{0,1\}^D$ of (P_1') can be mapped into a trajectory of a well chosen MDP. Consider the MDP with time horizon D, state

$$s_t = \left(x_{t-\Delta}, \dots, x_{t-1}, \sum_{t'=1}^{t-\Delta} x_{t'} \right),$$

action $a_t = x_t$ and reward function $r_t(s_t, a_t)$, where $r_t(s_t, a_t) = \infty$ if $t = D$ and $\sum_{t=1}^{D} x_t \neq m$ and

$$r_t(s_t, a_t) = x_t (1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}},$$

otherwise. By inspection, this problem is a MDP with finite horizon D, with initial state $(0, \dots, 0)$, deterministic transitions since s_{t+1} may be expressed as a deterministic function of the state-action pair (s_t, a_t). The state space is $\mathcal{S} = \{0,1\}^\Delta \times \{0, \dots, m\}$ and the action space is $\mathcal{A} = \{0,1\}$. Furthermore, the sum of rewards is:

$$\sum_{t=1}^{D} r_t(s_t, a_t) = \sum_{t=1}^{D} x_t (1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}},$$

which is precisely the objective function of problem (P_1'). Hence dynamic programming applied to this problem runs in time $\mathcal{O}(D|\mathcal{S}||\mathcal{A}|) = \mathcal{O}(mD2^\Delta)$ and memory $\mathcal{O}(|\mathcal{S}|) = \mathcal{O}(m2^\Delta)$ which yields the result. □

p	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
≤ 0.2	**1**	1	0	**0**	1	1	**0**	1	1	**0**	1	1
0.3	**1**	0	0	**1**	1	0	**1**	1	0	**1**	1	1
0.4	**1**	0	0	**1**	1	0	**1**	1	1	**1**	1	1
≥ 0.5	**1**	0	0	**1**	0	0	**1**	1	1	**1**	1	1

Table 4.1: Optimal policy for $D = 12$, $m = 8$, $\Delta = 3$ and $p \in [0, 1]$

In Table 4.1 we compute the optimal policy $\mathbf{x}^\star$ on a small example of $D = 12$, $m = 8$, $\Delta = 3$ and various values of p. The first remark is that $\mathbf{x}^\star$ does depend on p, so that one must estimate p in order to act optimally in general. Second, for all $p \geq 0.5$, the optimal policy has a particular form which we call a "semi-greedy" policy. Namely one first transmits at time instants $\{1, 1 + \Delta, 1 + 2\Delta, \dots\}$ (represented in red color on the table), and then one transmits in a long, uninterrupted burst until the deadline. Namely, we start by sending a packet every Δ time slots in order to receive the feedback and avoid spurious transmissions, and at the end we send a long burst in order to meet the constraint of $\sum_{t=1}^{D} x_t = m$. We study this phenomenon in greater details in the latter subsections.

We will use the following intermediate result in the latter sections, which provides a lower bound on the optimal value of (P_1). As a corollary, it shows the intuitive fact that, whenever $\Delta m \leq D$, the optimal policy simply consists in transmitting a packet every Δ time slots to make full use of the feedback about each transmission.

Proposition 4.2.2. *For all $D \geq 1$, $m \leq D$, $\Delta \geq 0$ and $p \in [0, 1]$ the optimal value of (P_1) admits the lower bound $g(\mathbf{x}^\star) \geq (1 - (1 - p)^m)/p$.*

Proof. Define $1 \leq t_1 < \cdots < t_m \leq D$ the slots where the optimal policy $\mathbf{x}^\star$ transmits, so that $x^\star_{t_k} = 1$ for $k = 1, \dots, m$. By definition we have $\sum_{t=1}^{t_{k-1}} x_t = k-1$, so that

$$\begin{aligned} g(\mathbf{x}^\star) &= \sum_{k=1}^{m} (1-p)^{\sum_{t=1}^{t_k - \Delta} x_t} \geq \sum_{k=1}^{m} (1-p)^{\sum_{t=1}^{t_{k-1}} x_t} \\ &= \sum_{k=0}^{m-1} (1-p)^{k-1} = \frac{1-(1-p)^m}{p}. \end{aligned}$$

□

Corollary 4.2.2. *Consider $D \geq 1$, $m \leq D/\Delta$ and $p \in [0, 1]$. Then the optimal policy $\mathbf{x}^\star$ is given by: $x_t = 1$ if $t = 1 + k\Delta$ for $1 \leq k \leq m$ and $x_t = 0$ otherwise.*

m	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
2	**1**	0	0	**1**	0	0	**0**	0	0	**0**	0	0
4	**1**	0	0	**1**	0	0	**1**	0	0	**1**	0	0
6	**1**	0	0	**1**	0	0	**1**	0	0	**1**	1	1
8	**1**	0	0	**1**	0	0	**1**	1	1	**1**	1	1
10	**1**	0	0	**1**	1	1	**1**	1	1	**1**	1	1
12	**1**	1	1	**1**	1	1	**1**	1	1	**1**	1	1

Table 4.2: Semi-greedy policy, $D = 12$, $\Delta = 3$ and $m \in \{2, 4, 6, 8, 9, 10\}$

Proof. One can check that the value of this policy equals $(1 - (1 - p)^m)/p$ and apply the previous result. □

4.2.3 The semi-greedy policy

As shown by previous results, the optimal policy depends on the success probability p and may be computed in reasonable time (but the computation is still an increasing function of the problem size). Two problems may arise in practical systems. First, in many practical systems, the success probability is initially unknown and must be estimated, which wastes time and radio resources if an accurate estimate is needed and second, in cases where very little computing power is available, computing the optimal policy on-the-fly might not be feasible. We now study the semi-greedy policy. This policy is illustrated in Table 4.2 on an example of $D = 12$ and $\Delta = 3$.

Definition 4.2.3. *The semi-greedy policy transmits one packet every Δ slots, and transmits the remaining packets as a burst in the last slots. Formally, this policy is $\widetilde{\mathbf{x}}$ with:*

$$\tilde{x}_i = \begin{cases} 1 & \textit{if} \quad (i-1) \mod \Delta = 0 \quad \textit{or} \quad i > i^\star \\ 0 & \textit{otherwise}. \end{cases}$$

with $i^\star = \Delta(k^\star) - a^\star$, $k^\star = \lfloor \frac{D-m}{\Delta-1} \rfloor + 1$ and $a^\star = 1 + m - D + k^\star(\Delta - 1)$.

In this section, we show that the semi-greedy is a good policy to use in practice due to two facts:

- If the success probability is high enough (p close enough to 1 as in many URLLC applications using robust modulation and coding schemes), the semi-greedy policy becomes optimal, so that no computation is needed, and one does not require an accurate estimate of p to act optimally.

- For large problems, i.e. $D \gg 1$, for any value of p, the semi-greedy policy is asymptotically optimal i.e. its approximation ratio goes to 1, and its value is asymptotically the value of the optimal policy.

Theorem 4.2.4. *Consider $D \geq 1$, $m \leq D$ and:*

$$p \geq p_\Delta \equiv \frac{\Delta - 3 + \sqrt{(\Delta-3)^2 + 4(\Delta-1)}}{2(\Delta-1)}.$$

Then the semi-greedy policy is optimal, i.e. $g(\tilde{\mathbf{x}}) = g(\mathbf{x}^\star)$.

Proof. We consider p and Δ fixed. We denote by $G(m, D)$ and $G^\star(m, D)$ the values of the semi-greedy and optimal policies respectively for any given value of (m, D). First we notice that, since $m \mapsto G^\star(m)$ is increasing and for $m = D$ the optimal policy transmits every slot we have:

$$\begin{aligned} G^\star(m, D) \leq G^\star(D, D) &= \Delta + \sum_{k=1}^{D-\Delta} (1-p)^k \\ &\leq \Delta + \sum_{k\geq 1} (1-p)^k = \Delta + \frac{1-p}{p}. \end{aligned}$$

Consider $m \leq D$ and $k = \max(1, m + \Delta - D))$. Let us prove that the optimal policy transmits exactly k times in interval $\{1, \ldots, \Delta\}$. If $k = \Delta$ this is trivial since this implies $m = D$. Assume that the optimal policy transmits at least $k+1$ times in interval $\{1, \ldots, \Delta\}$, then $G^\star(m, D) \geq k + 1$. On the other hand, an acceptable policy is to transmit k times in interval $\{1, \ldots, \Delta\}$, and then apply the optimal policy on time interval in interval $\{\Delta + 1, \ldots, D\}$, so that:

$$\begin{aligned} k + 1 \leq G^\star(m, D) &\leq k + (1-p) G^\star(m - k, D - \Delta) \\ &\leq k + (1-p)(\Delta + \frac{1-p}{p}). \end{aligned}$$

where we used the previous inequality. Subtracting k on both sides yields:

$$1 \leq (1-p)(\Delta + \frac{1-p}{p}) < (1-p_\Delta)(\Delta + \frac{1-p_\Delta}{p_\Delta}) = 1$$

since the r.h.s. is strictly decreasing and $p > p_\Delta$. This is a contradiction, and we have proven that the optimal policy is the policy which, for all m and D transmits $k = \max(1, m + \Delta - D))$ in the time interval $\{1, \ldots, \Delta\}$. Since the semi-greedy policy is the only policy which verifies this property, the semi-greedy policy is optimal. □

We now consider large problems so that $D \gg 1$, and we assume that $m = \rho D$ is proportional to D, with $\rho \in [0,1]$ a fixed number that we call the load, since it is the number of transmissions per unit of time. As announced, our second result is that the semi-greedy policy is asymptotically optimal for large problems, and that the approximation ratio vanishes exponentially fast, so that even for modest values of D, the semi-greedy policy is very close to optimal.

Theorem 4.2.5. *Consider $\rho \in [0,1]$. Then for any $p \in [0,1]$ $D \geq 1$ and $m = \rho D$, the approximation ratio of the semi-greedy policy is bounded as:*

$$1 \leq \frac{g(\widetilde{\mathbf{x}})}{g(\mathbf{x}^\star)} \leq \frac{1 + p\Delta(1-p)^{\lfloor \frac{D(1-\rho)}{\Delta-1} \rfloor - 1}}{1-(1-p)^{\rho D}}.$$

Furthermore, for any fixed values of ρ and p, the semi-greedy policy is asymptotically optimal:

$$\lim_{D\to\infty} \frac{g(\widetilde{\mathbf{x}})}{g(\mathbf{x}^\star)} = 1.$$

Proof. The semi greedy policy starts by at least $k^\star = \lfloor \frac{D(1-\rho)}{\Delta-1} \rfloor$ transmissions at times $\{1, 1+\Delta, \ldots, 1+(k^\star-1)\Delta\}$, and then $m - k^\star \leq D$ packets in a burst, so that:

$$\begin{aligned} g(\widetilde{\mathbf{x}}) &\leq \sum_{k=0}^{k^\star-1} (1-p)^k + (1-p)^{k^\star-1}(\Delta + \frac{1-p}{p}) \\ &= \frac{1-(1-p)^{k^\star}}{p} + (1-p)^{k^\star-1}(\Delta + \frac{1-p}{p}) \\ &= \frac{1}{p} + \Delta(1-p)^{k^\star-1}. \end{aligned}$$

where we used the same bound as in the previous theorem. From Proposition 4.2.2 and the fact that $m = \rho D$ we have the lower bound:

$$g(\mathbf{x}^\star) \geq \frac{1-(1-p)^m}{p}.$$

Combining both inequalities we obtain:

$$\frac{g(\widetilde{\mathbf{x}})}{g(\mathbf{x}^\star)} \leq \frac{1 + p\Delta(1-p)^{k^\star-1}}{1-(1-p)^m} = \frac{1 + p\Delta(1-p)^{\lfloor \frac{D(1-\rho)}{\Delta-1} \rfloor - 1}}{1-(1-p)^{\rho D}}.$$

□

Remark 4.2.1. *The case where m is fixed and D goes to infinity is trivial, in the sense that, as soon as $D \geq m\Delta$ the semi-greedy policy is optimal by Corollary 4.2.2.*

4.2.4 Numerical Experiments

We conclude this section by illustrating our results using numerical experiments. On Fig. 4.1 we compare the value of several policies for $D = 100$, $m = 60$ and a delay of $\Delta = 5$ slots. We represent the optimal policy $\mathbf{x}^\star$, the semi-greedy policy $\widetilde{\mathbf{x}}$ (denoted by "proposed policy"), a policy which transmits in the first m slots (denoted by "burst policy"), a policy which alternates between transmitting during $\beta = \lceil \frac{m}{D-m} \rceil$ slots and not transmitting during one slot (denoted by "burst each β policy"), and a policy which randomly selects m slots out of D to transmit (denoted by "random policy"). The semi-greedy policy is very close to the optimal policy, and outperforms all the other policies. As predicted by Theorem 4.2.4, for $p \geq p_\Delta$ the semi-greedy is exactly optimal, and interestingly, p_Δ seems to roughly match the threshold at which the optimal and the semi-greedy policies perform the same. We also observe that it seems to be better to transmit completely at random rather than send a long burst or regularly spaced small bursts.

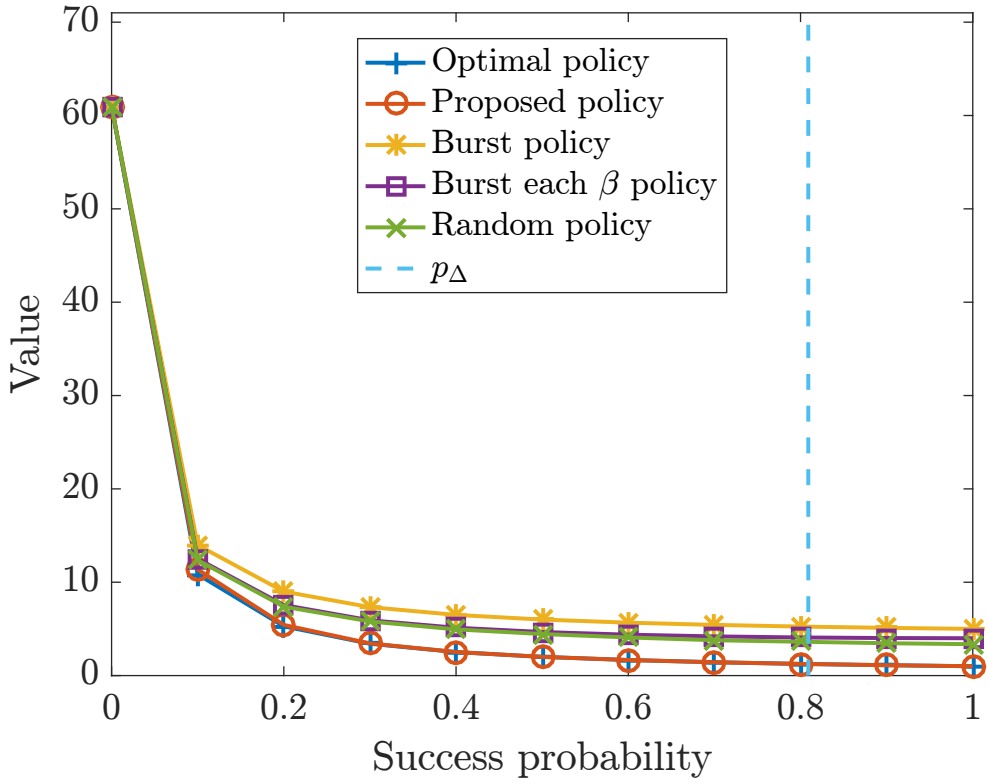


Figure 4.1: Value of the considered policies for $D = 100$, $m = 60$ and $\Delta = 5$

On Fig. 4.2 we plot the approximation ratio of each policy which is $\frac{g(\mathbf{x})}{g(\mathbf{x}^\star)}$ where

$\mathbf{x}$ is the considered policy. Not only do we observe that the approximation ratio is very close to 1 for the semi-greedy policy, but this is true for all values of p. Hence the semi-greedy policy is a suitable choice irrespective of the value of p.

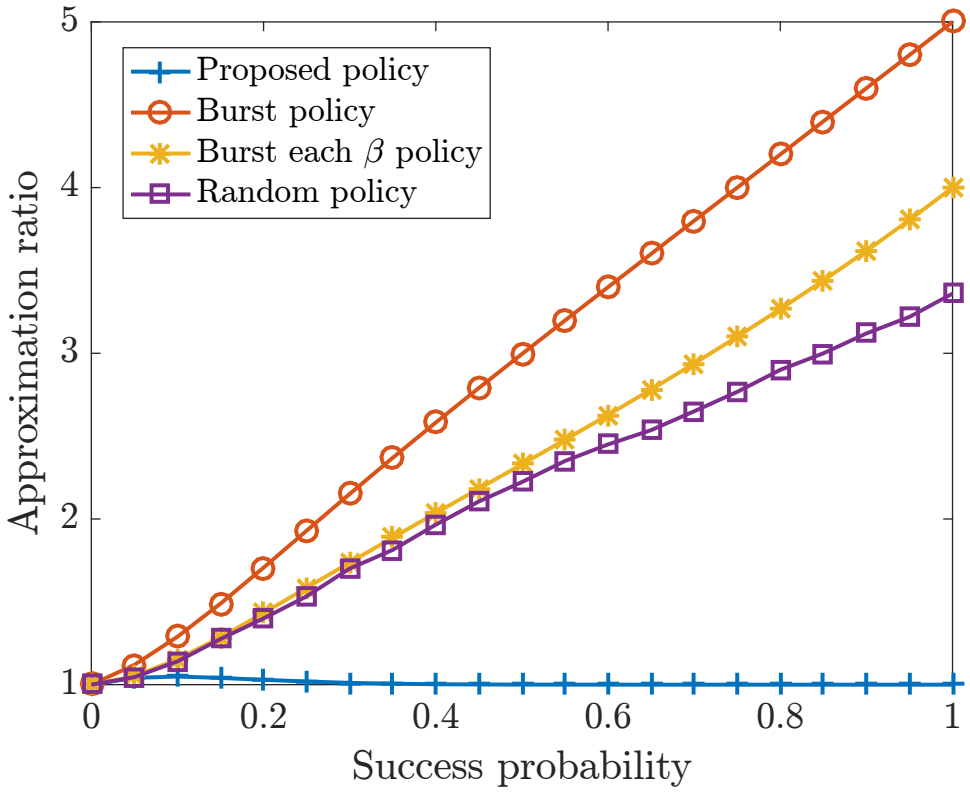


Figure 4.2: Approximation ratio of the considered policies for $D = 100$, $m = 60$ and $\Delta = 5$

In Fig. 4.3 we plot the approximation ratio of the semi-greedy policy as a function of p and m. The largest value of this ratio for any p and m is 1.08, so that the semi-greedy policy is both good and robust (as it does not depend on p), so that our numerical experiments agree with the asymptotic analysis of Theorem 4.2.5. It is also noted that for $m = 20$, we have $m\Delta = D$, so that the semi-greedy policy is optimal as predicted by Corollary 4.2.2. Once again, for $p \geq p_\Delta$ the semi-greedy policy is optimal.

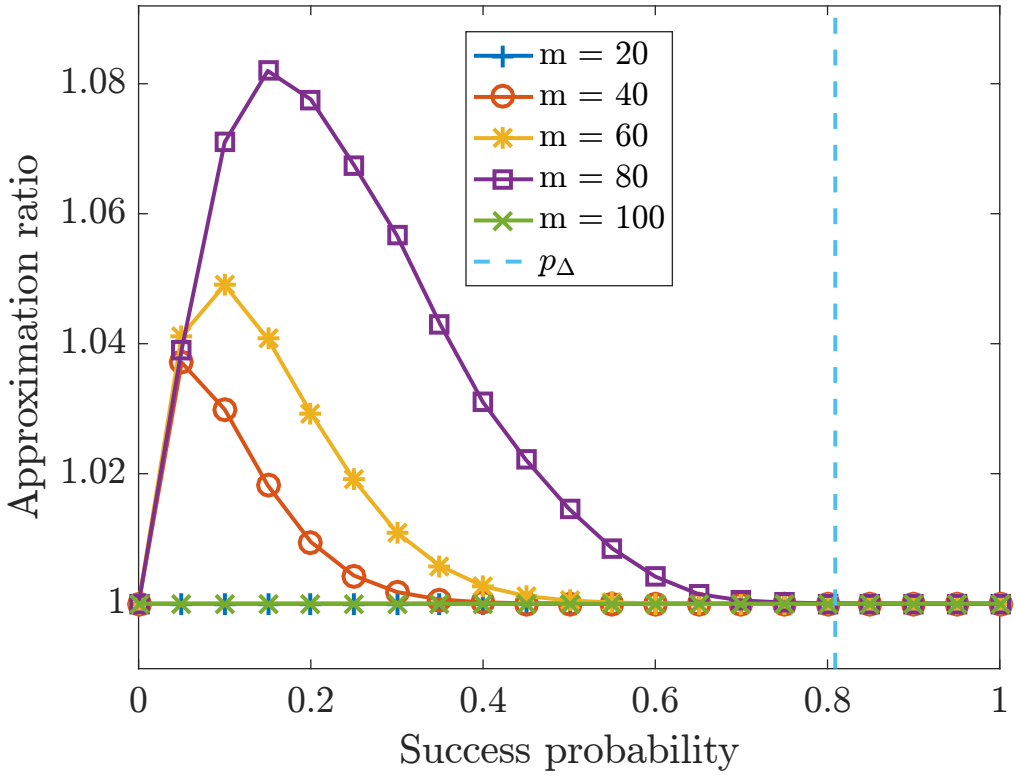


Figure 4.3: Approximation ratio of the semi greedy policy for $D = 100$ and $\Delta = 5$

4.2.5 Extension to the Multicast Case with Multiple Receivers

Link model. We now extend the previous case to a single UE transmitter whom attempts to transmit one packet to a multitude $N \geq 1$ of UE receivers, as in many URLLC use cases using D2D or V2V (e.g. a car trying to transmit an urgent packet to all cars in its vicinity). If the UE transmits at slot t, each UE receiver successfully decodes the packet with probability p. Let $Y_t \in \{0, \ldots, N\}$ be the number of UE receivers whom successfully decode the packet at time t, and not before. Then the packet is successfully delivered at slot t if and only if $\sum_{k=1}^{t} X_k Y_k = N$. If the packet is not completely delivered after D slots (the deadline) so that $\sum_{t=1}^{D} X_t Y_t < N$, it is considered lost. We assume that the success probability of a single link is $p \in [0, 1]$.

Feedback model. Whenever the UE transmits, it received feedback after a fixed delay of Δ time slots, indicating whether there is still someone who did not successfully decode the packet[2]. Denote by Z_t the feedback received at time t, then

[2]It is not practical to consider explicit feedback from each UE receiver in such a multicast scheme

we have

$$Z_t = \begin{cases} X_{t-\Delta} \mathbb{1}\{\sum_{k=1}^{t-\Delta} Y_k < N\} & \text{if } t \geq \Delta + 1, \\ 0 & \text{if } t \leq \Delta. \end{cases}$$

where $\mathbb{1}\{.\}$ is an indicator function equal to 1 if the event is true and to 0 otherwise.

Policies and objective. Our objective, denoted by problem (P_{1bis}), is to minimize the number of transmissions, subject to a reliability constraint, where the probability of not successfully decoding the packet is equal to a known threshold:

$$\begin{aligned} &\underset{\pi \in \Pi}{\text{minimize}} \sum_{t=1}^{D} \mathbb{E}(X_t^\pi) && (P_{1bis}) \\ &\text{subject to } \mathbb{P}\left(\sum_{t=1}^{D} X_t^\pi Y_t < N\right) = \epsilon = 1 - (1 - (1-p)^m)^N \end{aligned}$$

where $m \in \{0, \ldots, D\}$ depends on the reliability target ϵ as follows:

$$m = \left\lceil \frac{\ln(1 - (1-\epsilon)^{1/N})}{\ln(1-p)} \right\rceil$$

Proposition 4.2.3. *The optimization problem (P_1') gives also an optimal solution to problem (P_{1bis}), scaled to a larger number of transmissions m. An optimal solution X_t^π of (P_{1bis}) can thus be obtained from a $\mathbf{x}^\star$ solution of (P_1') by setting:*

$$X_t^\pi = x_t^\star \prod_{t'=1}^{t} (1 - Z_{t'}) \tag{4.1}$$

Proof. We first show that X_t^π given by equation (4.1) is optimal. Its expectation is given by:

$$\sum_{t=1}^{D} \mathbb{E}(X_t^\pi) = \sum_{t=1}^{D} x_t^\pi \mathbb{E}(\mathbb{1}\{N > \sum_{t'=1}^{t-\Delta} Y_{t'}\})$$

The event $A_t = N > \sum_{t'=1}^{t-\Delta} Y_{t'}$ occurs when at least one of the N UEs did not decode the packet before $t - \Delta$. Its probability is equal to $Pr[A_t] = [1 - (1 - (1 - p)^{\sum_{t'=1}^{t-\Delta} x_{t'}})^N]$. $\mathbb{1}\{A_t\}$ being a Bernoulli variable, we have:

$$\sum_{t=1}^{D} \mathbb{E}(X_t^\pi) = \sum_{t=1}^{D} x_t^\pi [1 - (1 - (1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}})^N])$$

where the UE transmitter may even ignore the number of UE receivers around him. The presence of at least a NACK is then considered as a global NACK.

Consider now the following optimization problem:

$$\underset{\mathbf{x}\in\{0,1\}^D}{\text{minimize}}\ g'(\mathbf{x}) \equiv \sum_{t=1}^{D} x_t[1-(1-(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}})^N] \qquad (P_1'')$$

$$\text{subject to } \sum_{t=1}^{D} x_t = m.$$

As $\sum_{t=1}^{D} x_t = m$, we have $g'(\mathbf{x}) = m - \sum_{t=1}^{D} x_t[(1-(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}})]^N$
Recalling that $x_t \in \{0,1\}$, minimizing $g'(\mathbf{x})$ is equivalent to maximizing:

$$\sum_{t=1}^{D} x_t(1-(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}}) = m - \sum_{t=1}^{D} x_t(1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}} \qquad (4.2)$$

i.e. to minimizing $g(\mathbf{x})$, which concludes the proof. □

4.3 Multiple Transmitters

Throughout this section we consider the model with multiple UE transmitters described in Subsection 4.1.2, we design policies for problem (P_2), and provide techniques to analyze their performance.

4.3.1 Open Loop Policies

Policies. As mentioned before, we solely consider policies which are distributed and symmetrical, so that any UE transmitter only relies on the feedback it has seen, and all UE transmitters behave in the same manner. At time t, we write $A_{t,i} = 1$ if i has transmitted successfully at least once before t and $A_{t,i} = 0$ otherwise. Hence:

$$A_{t,i} = \mathbb{1}\left\{\sum_{t'=1}^{t-1}\sum_{j\neq i} X_{t',i}(1-X_{t',j}) > 0\right\},$$

If i transmits successfully for the first time at time t, then it is informed of this fact at time $t+\Delta$ through available feedback. Define $B_{t,i} = A_{t-\Delta,i}$, then we have $B_{t,i} = 1$ if and only if i has been informed that it has already successfully transmitted at least once before t. We say that i is active if $B_{t,i} = 0$ and inactive otherwise. Clearly $B_{t,i}$ can be observed by UE transmitter i at time t, and i should transmit only if it is active, since it needs to convey only one packet to the UE receiver. Denote by $N_t = \sum_{i=1}^{n} A_{t,i}$ the number of UEs that have successfully

transmitted a packet. We recall that the goal is to maximize the success rate $\frac{1}{n}\mathbb{E}(N_t)$. We define $q_t = \mathbb{E}(A_{t,i})$ the probability that i has successfully transmitted a packet before t, where q_t does not depend on i by symmetry and we note that the success rate is $\frac{1}{n}\mathbb{E}(N_t) = q_t$. We aim at maximizing q_t.

Open loop policies. For simplicity and practicality, we will consider a family of policies called open loop policies, where the transmission probability depends solely on the state of i at time t (active or not active) and time t.

Definition 4.3.1. *An open loop policy is a policy such that, at time t, each UE i transmits with probability p_t if it is active and 0 otherwise, independently of other UE transmitters, where $(p_t)_{t=1,\dots,D}$ is some known sequence.*

Time dependent policies. Let us first consider the case where there is no delay, so that $\Delta = 1$. We will extend our results to the general case in Subsection 4.3.3. We now aim at finding a good open loop policy. The simplest open loop policy is the ALOHA policy [111], where the transmission probability $p_t = p$ is constant. It is well known that, if ALOHA is applied to a system with n active UE transmitters with infinitely many packets to transmit, the choice maximizing the throughput is $p = \frac{1}{n}$. In our setting the number of active users $n - N_t$ is decreasing with time so that it would be natural to choose $p_t = \frac{1}{n-N_t}$. However this is not feasible, since N_t is unknown to UE transmitter i. So p_t should be a well chosen increasing function of time. Our aim is to maximize q_t, whose evolution equation is:

$$q_{t+1} = q_t + (1 - q_t)p_t\mathbb{E}\left((1 - p_t)^{n-N_t-1}\right).$$

This is due to the fact that there are $n - N_t$ active UE transmitters at time t, and that i has transmitted successfully before $t + 1$ if either it has transmitted successfully before time t (probability q_t) or if it is active at time t and transmits at t (probability $p_t(1 - q_t)$), and no other active UEs attempt to transmit at this time (probability $\mathbb{E}((1-p_t)^{n-N_t-1})$). Differentiating, we get that to maximize q_{t+1} knowing q_t, p_t should be such that:

$$\mathbb{E}\left((1 - p_t(n - N_t))(1 - p_t)^{n-N_t-1}\right) = 0$$

Therefore, choosing p_t requires to know the distribution of N_t, and there seems no obvious way to solve the problem.

4.3.2 A Mean Field Approach

Correlation between UE transmitters. The difficulty is that although $N_t = \sum_{i=1}^{n} A_{t,i}$ is a sum of n random variables with mean q_t, those random variables are not independent since UE transmitters interact with each other, and it is unclear what

is the distribution of N_t. In order to solve this problem we propose to use a *mean field* approach: we approximate the system by an alternative system in which UE transmitters are independent of each other, so that $A_{t,1}, \ldots, A_{t,n}$ are i.i.d. Of course, while this is an approximation, this type of mean field approach tends to be accurate in large systems with many UE transmitters $n \gg 1$, see for instance the seminal [112].

Evolution equations. Under the mean field approximation $N_t \sim \text{Binomial}(n, q_t)$ and:

$$q_{t+1} = q_t + (1 - q_t)p_t \left(1 - (1 - q_t)p_t\right)^{n-1}.$$

Define $a_n = (1 - \frac{1}{n})^{n-1}$. We now state Proposition 4.3.1 showing that under the mean field approximation, the system can be completely characterized: we can compute the optimal policy exactly, as well as the success rate. The result shows that the system goes through two phases: in the first phase the proportion of UE transmitters whom have succeeded q_t increases linearly and p_t increases in order to stay close to $\frac{1}{n-N_t}$. In the second phase, $p_t = 1$ is constant. It is also noticed that the time at which we switch from the first to the second phase $t = \frac{n-1}{a_n}$ is roughly the time required for $n - 1$ UE transmitters to successfully transmit.

Proposition 4.3.1. *Under the mean field assumption, the optimal choice of p_t is given by:*

$$p_t = \frac{1}{\max(n - (t-1)a_n, 1)}.$$

Under this policy we have that:

$$q_t = \frac{a_n(t-1)}{n} \text{ if } t \leq \frac{n-1}{a_n} \text{ and}$$
$$q_D \geq \min\left(1 - \frac{1}{n}, \frac{Da_n}{n}\right).$$

Remark 4.3.1. *Consider n and D large with $\frac{n}{D} \leq 1/e$. Then the success rate of the policy, q_D, tends to* 1.

Proof. Recall that the evolution equation for q_t is:

$$q_{t+1} = q_t + (1 - q_t)p_t \left(1 - (1 - q_t)p_t\right)^{n-1}.$$

Maximizing the r.h.s. with respect to $p_t \in [0, 1]$ we get:

$$p_t = \frac{1}{\max(n - (t-1)a_n, 1)}.$$

If $q_t \leq 1 - \frac{1}{n}$ we get, by recursion:

$$q_{t+1} = q_t + \frac{a_n}{n} = q_1 + \frac{ta_n}{n} = \frac{ta_n}{n}.$$

for all $t \leq \bar{t} \equiv (n-1)/a_n$ which is the first statement. If $Da_n \leq n-1$, we have $\bar{t} > D$ so that $q_D = Da_n/n$. Otherwise, $\bar{t} \leq D$ so that $q_D \geq q_{\bar{t}} \geq 1 - 1/n$. Putting both cases together yields the result. □

4.3.3 Extension to the Delayed Case

Now consider the delayed case. Consider D and n large, and Δ small with respect to n as done in the previous sections. We argue that in this setting, the system is equivalent to the system without delays that we have studied previously. We do not give a formal proof and we propose a heuristic justification. In the delayed case, the proportion of active users is $1-q_t = \frac{1}{n}\sum_{i=1}^{n}(1-A_{t-\Delta,i})$ and the proportion of users that would be active in a non delayed system is $1-q'_t = \frac{1}{n}\sum_{i=1}^{n}(1-A_{t,i})$. At most one UE may transmit successfully at a time hence $0 \leq \sum_{i=1}^{n}(A_{t,i} - A_{t-\Delta,i}) \leq \Delta$. Therefore $|q_t - q'_t| \leq \frac{\Delta}{n}$. Hence, in the regime where $\Delta \ll n$, the proportion of active users with and without delays is the same, and the mean field model should be the same in both cases when there are a large number of users. So we may apply the same policy as in the non-delayed case and the performance should be similar. This is confirmed by our numerical experiments.

4.3.4 Numerical Experiments

We conclude by some numerical experiments. We define the system load $\alpha = \frac{n}{D} \in [0, 1]$. On Figs. 4.4, 4.5 and 4.6 we compare the value of the success rate $\frac{\mathbb{E}(N_t)}{n}$ computed from simulations (denoted by "S") to the value predicted by the mean field approximation (denoted by "MV"). The y-axis is the success rate and x-axis is the normalized time $\frac{t}{D}$. We draw four conclusions: (i) the mean field approximation is rather accurate even for a modest value of n, (ii) its accuracy increases with n, (iii) the optimal policy derived under the mean field approximation performs quite well in the original system, and (iv) the impact of the delay is very small as long as it is much smaller than the number of users $\Delta \ll n$. Hence the mean field provides a good method for solving an otherwise intractable problem, even when delays are present.

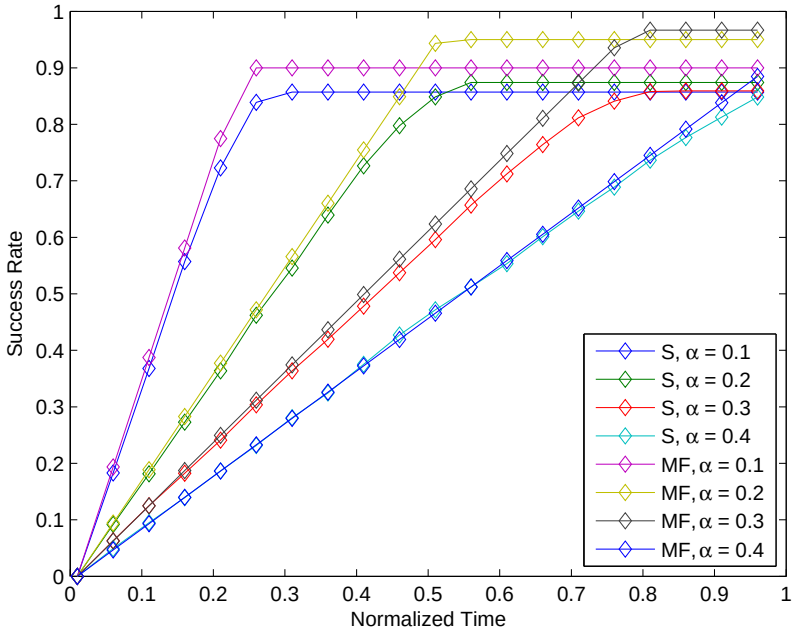


Figure 4.4: Mean field approximation vs simulation for $D = 10^2$ and $\Delta = 1$.

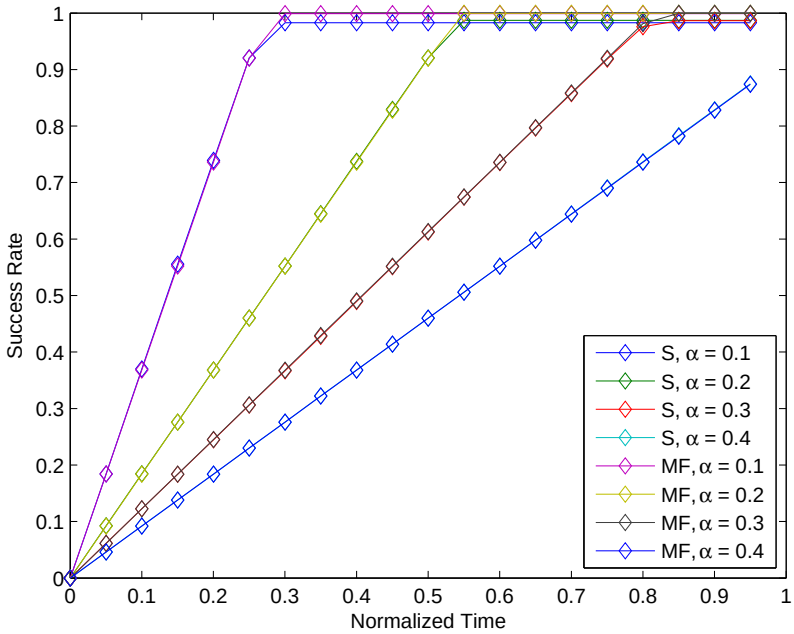


Figure 4.5: Mean field approximation vs simulation for $D = 10^4$ and $\Delta = 1$.

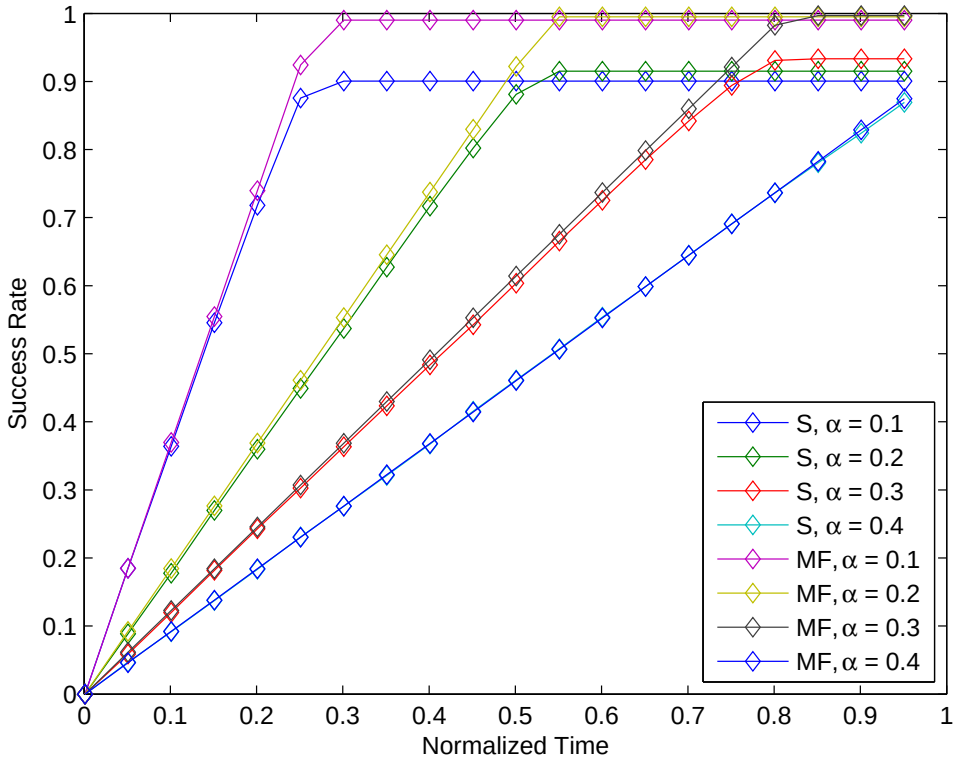


Figure 4.6: Mean field approximation vs simulation for $D = 10^3$, $\Delta = 5$

In Fig. 4.7, inspired by the form of the optimal policy in the mean field approximation, we study a parameterized family of policies with transmission probability:

$$p_t = \frac{1}{\max(n - c(t-1), 1)}$$

where $c \in [0, 1]$ is the policy parameter. For $c = 0$ this is simply ALOHA, and for $c = a_n$ we obtain the optimal mean field policy. We plot the success rate of this policy as a function of the load α and parameter c. Not only does this policy perform very well for c well chosen, but this also seems to suggest that $c \approx \alpha$ seems to be a good choice whenever $\alpha \leq 1/e$.

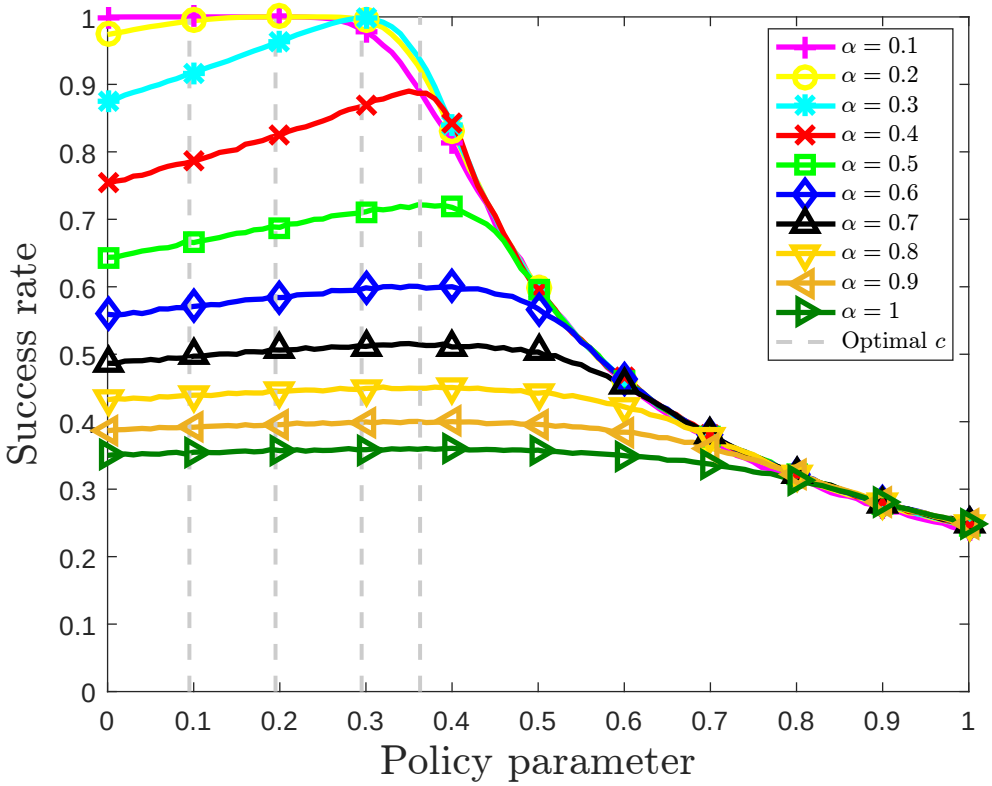


Figure 4.7: Optimal value of policy parameter c for $D = 200$ and $\Delta = 5$

Finally in Fig. 4.8 we compare the success rate of various policies, including ALOHA, where $p_t = \frac{1}{n}$ is constant, our "proposed policy" which is the policy studied in Fig. 4.7 using the optimal value of c, CSMA, which is an adaptive version of ALOHA where the transmission probability is $p_t = 2^{-k_{t,i}}$, where $k_{t,i}$ is the number of collisions seen by i up to time t (see [112]), and CSMA+, an adaptive version of the "proposed policy" in which the transmission probability is:

$$p_t = \frac{1}{\max(b^{-k_{t,i}} - c(t-1), 1)},$$

and where we optimize over $b \in [0,1]$ and $c \in [0,1]$. It is noted that the last two policies are not open loop policies (they take into account the number of collisions), and are adaptive versions of ALOHA and the "proposed policy" where one does not need to know n in advance. The best performance is obtained by our proposed policy, and as long as $\alpha \leq \frac{1}{e}$, its success rate is very close to 1.

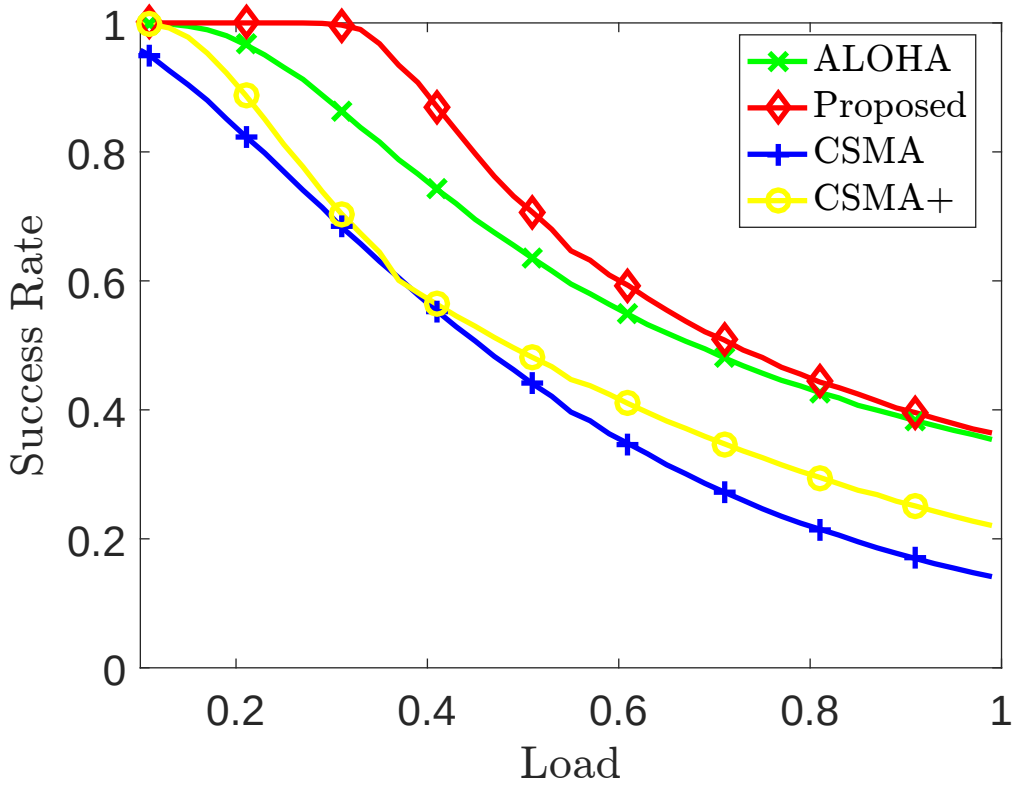


Figure 4.8: Success rate of the various policies for $D = 200$ and $\Delta = 5$

4.4 Conclusion

In this chapter we have studied the problem of designing HARQ schemes in systems which have both latency and reliability constraints. This problem is motivated by the design of URLLC services in future networks, in particular for V2X communications, e-health tele-surgery or industry 4.0. In the case of a single UE transmitter in a noisy environment, we derive the optimal policy and prove that the semi-greedy policy is a good choice in practice. For multiple UE transmitters, we propose policies with good performance and a mean field analysis to predict their performance.

4.5 Appendix

We recall below the list of parameters used along the chapter:

Table 4.3: List of parameters in Chapter 4

D	deadline constraint
Δ	feedback delay
m	number of retransmissions
n	number of UE transmitters
N	number of UE receivers
ϵ	reliability target
$X_{t,i} = \{0,1\}$	transmission decision at time t by UE i
$X^{\pi}_{t,i} = \{0,1\}$	transmission decision at time t by UE i for policy π
$Y_t = \{0,\dots,N\}$	number of successful receptions at time t
$Z_{t,i} = \{0,1\}$	received feedback at time t by UE i
$A_{t,i} = \{0,1\}$	UE i has transmitted successfully/unsuccessfully at least once before time t
$B_{t,i} = A_{t-\Delta,i}$	UE i has transmitted successfully/unsuccessfully at least once before time $t-\Delta$
$N_{t,i} = \sum_{i=1}^{n} A_{t,i}$	number of transmissions that have been successful before time t
Π	set of all policies
$\pi^{\star}$	optimal policy
$\mathbf{x} \in \{0,1\}^D$	transmission sequence
$\mathbf{x}^{\star} \in \{0,1\}^D$	optimal solution
$\widetilde{\mathbf{x}} \in \{0,1\}^D$	semi-greedy solution
$g(\mathbf{x}) \equiv \sum_{t=1}^{D} x_t (1-p)^{\sum_{t'=1}^{t-\Delta} x_{t'}}$	function to minimize over $\mathbf{x}$ in (P_1')
s_t	state t
$\mathcal{S}$	state space
a_t	action t
$\mathcal{A}$	action space
$r_t(s_t, a_t)$	reward function
p	transmission success probability
p_{Δ}	threshold probability such that $\widetilde{\mathbf{x}}$ is optimal
p_t	transmission probability at time t
q_t	probability that a UE has successfully transmitted before time t

Chapter 5

Device-to-Device aided Group Localization

In this chapter we provide a cooperative high accurate mapping for smart devices without any help of the network infrastructure. For this, we suppose that our devices are able to directly communicate using D2D technology. We propose a simple algorithm that uses as input a limited number of the distances between pairs of devices provided by ranging techniques like ToA, performed by using D2D between the devices. Four different channel models are studied in this chapter: LOS and NLOS for indoors and outdoors environments. The results show significant improvement in terms of accuracy, energy consumption, radio resource utilization and device complexity compared to already existing methods. Such localization algorithm is clearly suitable for services in smart cities within IoT context, but not limited to it. Still in the context of 5G, we observe an increasing tendency of local and instantaneous services for scenarios like bus stops, concerts, stadium events, city meeting points and other locations with dense mobile traffic. The high number of devices in such locations forms groups of devices that we may want to locate and track. As we evolve through a dynamic Radio Access Network (RAN), identifying the dense traffic areas in real time and providing a good quality of service arises as a key challenge to be addressed. Aided by a large number of antennas at the transmission, the BS can focus the energy on a group of UEs, increasing the network capacity through a virtual densification, therefore, avoiding the deployment cost of new infrastructure. A direct application of the proposed localization algorithm to this last scenario, also called Virtual Small Cells (VSCs), is detailed in this chapter.

In summary, our contributions are outlined below:

1) In this third work we propose a dynamic cooperative equipment mapping using Ultra-WideBand (UWB) signals, where map accuracy is studied in

LOS and NLOS channel in indoor and outdoor environments providing cm reliability.

2) An application of these UEs using UWB signal transmission has been applied to coverage scenario, named VSC, useful in 5G services like ultra-low cost global coverage, smart cities or industry 4.0. More specifically, the chapter highlights the importance of hotspot location knowledge to efficiently implement VSCs.

3) Due to hotspot location impact, the paper proposes to study and compare our cooperative group localization techniques with other common method based on GPS.

This chapter is organized as follows: Section 5.1 provides discussions on the existing localization systems and ranging techniques that can be used for our problematic. Section 5.2 describes our general approach for devices mapping, followed by a simple algorithm to improve its accuracy in Section 5.3. The simulation results are presented in Section 5.4. Section 5.5 highlights an important scenario with direct application of our device mapping algorithm, namely VSCs and finally Section 5.6 summarizes our conclusions.

5.1 Localization and Ranging Techniques

5.1.1 Localization Systems

In the literature several localization systems exist; GPS is the most well-known and used localization technique. However GPS presents a weakness for indoors localization due to the difficulties for its waves to penetrate the walls [113] [114]. Hence, GPS is limited to outdoor applications. A clustering method has already been developed in [115] in order to find groups of UEs of the size of a picocell (around 40 m) using K-means algorithm and GPS coordinates but its performance is limited by the GPS accuracy.

On the radio side, there are several promising localization technologies for indoors and outdoors. However, as we want to achieve an accuracy of centimeters, some technologies such as Radio Frequency Identification (RFID), Blue-tooth or WLAN fails to meet our standards [116]. The ones satisfying our accuracy requirements are ultrasound, InfraRed, Zig-bee and UWB. Ultrasound offers an accuracy of 10 cm [116] but the waves are impracticable in confined and large areas because, as for InfraRed, it cannot penetrate obstacles [117]. Zig-bee is weak to interference of other systems that use the same frequency (large bands systems) and may disturb radio communications if the communication takes more than few

milliseconds [118]. And finally UWB emits short low energy pulses spread in a very large bandwidth. Therefore UWB can achieve high immunity against interference and fading, low energy consumption [106] and very high accuracy thanks to its wide frequency band [119] as it is indicated by CRLB [107]. The only considerable disadvantage of UWB is its coverage range which can be solved thanks to a well-organized routing topology between the smart devices. These UWB distinct features offer UWB the best advantages of robustness, accuracy and energy consumption compared to other technologies. Furthermore, UWB is part of the IEEE 802.15 working group which specifies WPAN standards [17]. In that sense, we may expect a future use of UWB for D2D communications[1].

5.1.2 Ranging Techniques

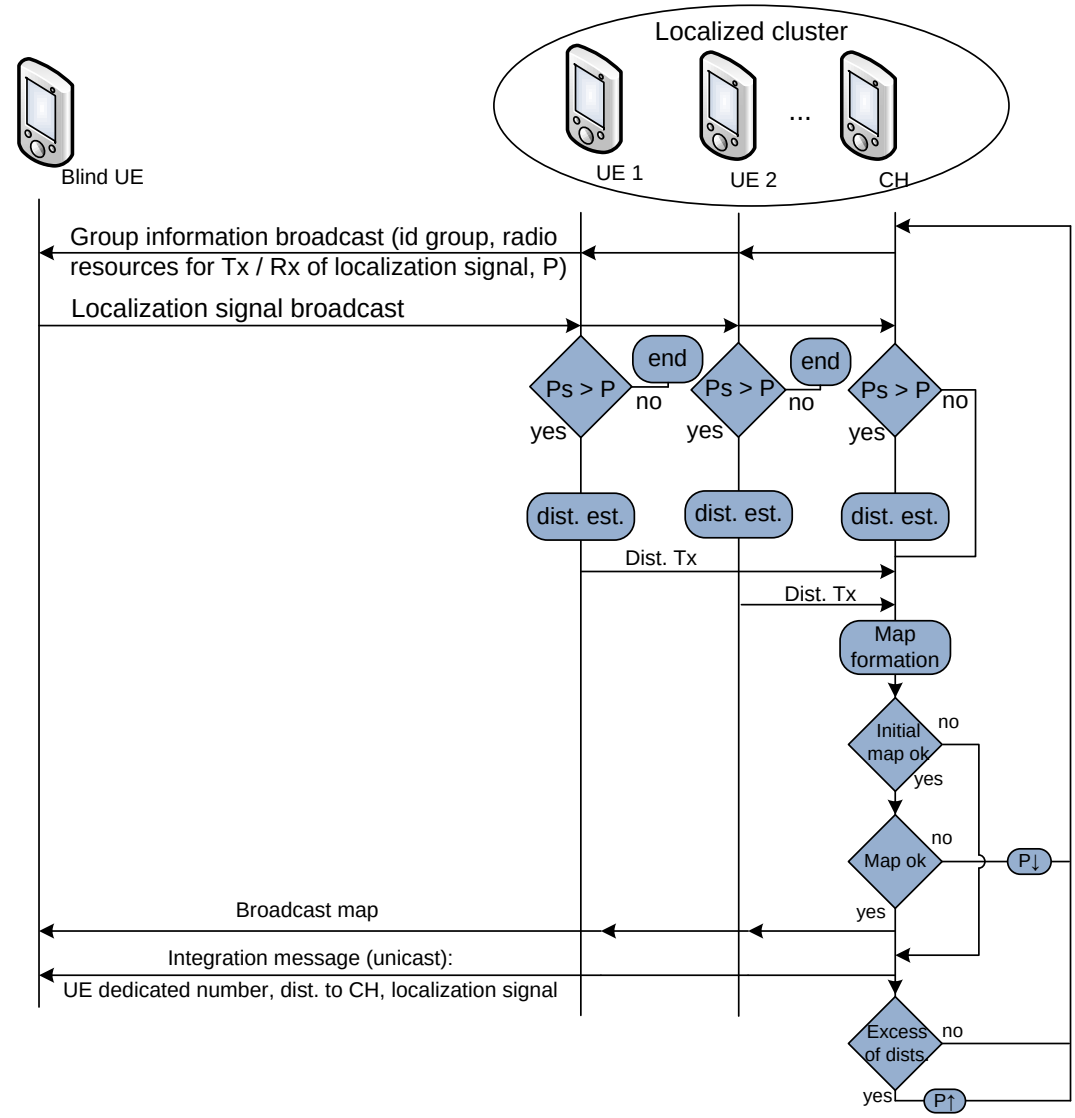
Given that we choose UWB as localization system, we will see the different techniques to estimate the distance between a pair of smart devices. AoA, ToA and Received Signal Strength (RSS) are the most popular ones. AoA requires multiple antennas to estimate the angle and therefore the localization will require large and expensive devices. This will cancel one of the main features of UWB which is its low cost equipments [120]. Another angle estimation challenge resides in the requirement of clear LOS between transmitter and receiver. With AoA the slightest angle estimation error impacts negatively the distance estimation. RSS is a measurement of the power received by a radio signal receiver. However these measurements are easily affected by the obstacles, thus RSS are unpredictable and their measurements are not accurate. ToA is a one way time difference between the moment the receiver detects the transmitters' signal and the time when the transmitter sends the signal. ToA offers the best accuracy for UWB as CRLB had demonstrated. This technique requires perfect time synchronization between different transmitters and receivers to precisely estimate the distance. The major ToA challenge is the relative clock drift between different nodes. Synchronization problem has been addressed by several works and different promising solutions have been proposed [12]. Furthermore, due to the large bandwidth of UWB signal, multi-path components are often resolvable without the use of complex algorithms [121]. And due to existing high performance ToA estimation for UWB in NLOS [122, 123, 124], we will focus on this ranging technique. This mapping UE aided by D2D connection could be useful for applications like: environmental studies (polluted zones, coal mining etc.), social networking, home/packet warehouse application, animals localization in a zoo, connected cars or deep indoor localization, in addition to VSCs.

[1]Please note that our method is not limited to UWB, actually with the ongoing tendency of increasing the bandwidth within 5G, and depending on de required services accuracy, we could easily apply our method to cellular 5G networks.

5.2 Equipment Mapping using Direct Link

The objective of our algorithm is to create a map of UEs by using only the distances between some of them as inputs and without anchor nodes, as in [125]. Those distances input between the UEs could be estimated by the ranging techniques presented in Section 5.1. Our approach uses ToA for the reasons previously mentioned. The UEs that are able to communicate are the ones creating the map, and possibly several groups of UEs will be observed in this map.

First of all, a Cluster Head (CH) is chosen according to the desired strategy. The CH could be a UE with a specific processor that is able to generate the computation and the treatment of the algorithm map. Note that the CH could also be the BS, or a specific UE within a group (for example the fireman leader in a group of firemen), or just a device chosen randomly within the group of devices to be mapped. The CH could be the first UE launching the algorithm map. In a first step, the CH will broadcast an information group message to the neighbors UEs including a group ID, radio resources for transmission and reception of localization signal and also a threshold P of power limit which is intended for the UEs already localized if any. Utility of P will be further explained. Once a neighbor UE receives the information group message previously stated, if this UE is interested by joining the group, he will broadcast back a positive response to the UEs already localized by using the radio resources provided by the CH. As a result, and by using this broadcast signal, the already localized UEs will be able to evaluate the distance between the transmitter and the receiver and will transmit this information to the CH. In order to have a good estimation of the distance, we introduce a received power threshold P added to the already accurate UWB method used for distance estimation, as presented in the introduction. Indeed, if a localized UE receive a signal from an unlocalized UE k, also called blind UE, with a received power $P_k < P$, the distance estimation procedure will not be lunch. The value of P depends on the required accuracy needed by the application, and its objective is to reduce the number of useful measures. High accuracy implies small value of P and as a consequence more measures, more energy consumption and more radio resources. Hence only reliable signals will be considered for distance estimation and the combination of the given distances will enable us to find the location of the UEs. Finally, if a UE is added to the map, a broadcast of the map is possible and an integration message is sent by the CH to the new UE. The previous description is represented in Fig. 5.1. We will describe in details all the steps of this method in the next subsections.



Ps: received power signal strenght
P: threshold learning parameter based on signal power
In the case of several UE detection, UE's dedicated number is based on distance with CH so as to differentiate them

Figure 5.1: General Map Steps

5.2.1 Initial Map

During the initialization step of the map, as we have limited distance knowledge of the neighbors UEs, any UE is able to integrate our map as long as the threshold received P power is satisfied. The initialization map is the smallest possible map that we are able to form. We can achieve at most a combination C_n^2 of distances

for n UEs. Actually this smallest map is composed of four UEs and needs six distances between them. This is observed in Proposition 5.2.1.

Proposition 5.2.1. *With only $n = 4$ UEs in Two-dimension (2D), and $n = 6$ in Three-dimension (3D), and the knowledge of $C_n^2 = 6$ distances, $C_n^2 = 15$ respectively, named $d_{ij} \in \mathbb{R}$ between UE i and j ($\forall\{i,j\}_{i=1,j>i}^{n}$), we can map the n UEs in an orthonormal coordinate system where the origin corresponds to the CH or any other UE.*

Proof. Consider n UEs and d_{ij} distances between i and j:
- UE 1, (x_1, y_1), is connected to $n-1$ other UEs,
- UE 2 is also connected to the same number of UEs, but only $n-2$ distance values are new as $d_{1,2} = d_{2,1}$; $d_{2,i}$ $\forall i \in [3, n]$,
etc.
- UE $n-1$ gives us one new distance $d_{n-1,n}$.
We have the knowledge of $\sum_{i=1}^{n-1}(n-i) = \frac{n(n-1)}{2}$ distance values, i.e. $C_n^2 = \frac{n(n-1)!}{2(n-2)!}$ combination of distances.
Also, the location of all the UEs is unknown unless for the selected UE o, $(x_o, y_o) = (0, 0)$. So finally, we do not know the $2(n-1)$ remaining coordinates.
i.e. we need at least $\frac{n(n-1)}{2} = 2(n-1)$, $n = 4$. □

Within those four UEs we have one CH. The CH is aware of the six distances between them, and each distance provides an equation of the type $d_{ij} = \sqrt{(y_j - y_i)^2 + (x_j - x_i)^2}$ between UE i and j, with four unknown values (two UEs in 2 dimensions). We then have six equations of this type for the case of four UEs (eight unknown values). Actually as the wanted map is relative to the CH, we can consider that the CH coordinates are at the origin of the map and hence are $(0, 0)$. So we finally have six equations for six unknown values, which is solvable. If the system of six equations provides a solution, i.e. if the distances are well estimated, we obtain our initial map. In the case of three dimensions, we can demonstrate that the smallest map is composed of six UEs and we will need fifteen distances.

Remark 5.2.1. *When creating this map, we observe infinite solutions, but all the solutions will provide the same map with some rotation between 0 and 2π. Uniqueness of the map can be achieved by the knowledge of the coordinate of one UE (different than the CH) in 2D and one of the coordinates of two UEs in 3D.*

Proof.

$$\begin{cases} UE\ 1 \rightarrow & \begin{cases} 1 & = & \sqrt{(y_2)^2 + (x_2)^2} \\ \sqrt{2} & = & \sqrt{(y_3)^2 + (x_3)^2} \\ \sqrt{5} & = & \sqrt{(y_4)^2 + (x_4)^2} \end{cases} \\ UE\ 2 \rightarrow & \begin{cases} 1 & = & \sqrt{(y_3 - y_2)^2 + (x_3 - x_2)^2} \\ \sqrt{2} & = & \sqrt{(y_4 - y_2)^2 + (x_4 - x_2)^2} \end{cases} \\ UE\ 3 \rightarrow & \begin{cases} 1 & = & \sqrt{(y_4 - y_3)^2 + (x_4 - x_3)^2} \end{cases} \end{cases}$$

In this system we consider $n = 4$ UEs and $C^2_{n=4} = 6$ distances. Let UE 1 be at the origin of our system map, $(x_1, y_1) = (0, 0)$. Let $d_{1,2} = 1$, $d_{1,3} = \sqrt{2}$, $d_{1,4} = \sqrt{5}$, $d_{2,3} = 1$, $d_{2,4} = \sqrt{2}$ and $d_{3,4} = 1$.

The couples solution of this system is given by:

$$UE\ 2 = \begin{pmatrix} x_2, & -\sqrt{1-x_2^2} \\ x_2, & \sqrt{1-x_2^2} \\ x_2, & -\sqrt{1-x_2^2} \\ x_2, & \sqrt{1-x_2^2} \end{pmatrix}; UE\ 3 = \begin{pmatrix} x_2 - \sqrt{1-x_2^2}, & -x_2 - \sqrt{1-x_2^2} \\ x_2 - \sqrt{1-x_2^2}, & x_2 + \sqrt{1-x_2^2} \\ x_2 + \sqrt{1-x_2^2}, & x_2 - \sqrt{1-x_2^2} \\ x_2 + \sqrt{1-x_2^2}, & -x_2 + \sqrt{1-x_2^2} \end{pmatrix};$$

$$UE\ 4 = \begin{pmatrix} 2x_2 - \sqrt{1-x_2^2}, & -x_2 - 2\sqrt{1-x_2^2} \\ 2x_2 - \sqrt{1-x_2^2}, & x_2 + 2\sqrt{1-x_2^2} \\ 2x_2 + \sqrt{1-x_2^2}, & x_2 - 2\sqrt{1-x_2^2} \\ 2x_2 + \sqrt{1-x_2^2}, & -x_2 + 2\sqrt{1-x_2^2} \end{pmatrix};$$

We observe infinite solution providing the same map with some rotation. By fixing x_2, we obtain a unique map. □

Once the initial map is formed, this means that we have a map of at least four UEs.

5.2.2 General Map

In order to add a blind UE k to our map we can simply use a trilateration method where we need the distances between the new UE and three other UEs (h, i and j) already localized. The blind UE k is located at the intersection of the three circles of center h, i and j and radius d_{hk}, d_{ik} and d_{jk} respectively. The three distances used to localize the blind UE k correspond to the three most accurate selected thanks to the received power. If the threshold power P is too high such that the CH collects at least three signals, P will be reduced at the next iteration.

Proposition 5.2.2. *In 2D with $n > 4$, and in 3D with $n > 6$, a map can be created with only $6 + 3(n-4)(< C^2_n)$ distances, $15 + 4(n-6)$ respectively.*

Proof.
Initialization: $i = 5$
From Proposition 5.2.1, for $i = 4$ UEs, we need $C^2_i = 6$ distance values. If we add a UE to this map (i = 5), we will need 3 more distances based on triangulation. The maximum distances for $i = 5$ is $C^2_5 = 10$ distances, and $6+3 = 6+3(5-4) = 9 < 10$
Step $i = n$: $6 + 3(n-4) < C^2_n = \frac{n(n-1)}{2}$
Step $i = n + 1$:

$$6 + 3(n-4) + 3 = 6 + 3((n+1) - 4) < C_n^2 + n = C_{n+1}^2 \qquad \square$$

Proposition 5.2.2 means that we do not need all the distances between all the UEs in order to form a map. In order to add a blind UE to our map we will need to solve a system of three equations of the type $d_{ki} = \sqrt{(y_i - y_k)^2 + (x_i - x_k)^2}$ where k is the index of the blind UE and i represents three different localized UEs. Also for good accuracy of the system, distances estimation is very important as even with a small obstacle between two UEs, the distances estimation will be impacted. In order to fix this, we introduce a measurement incertitude threshold e (in meter) which is further detailed.

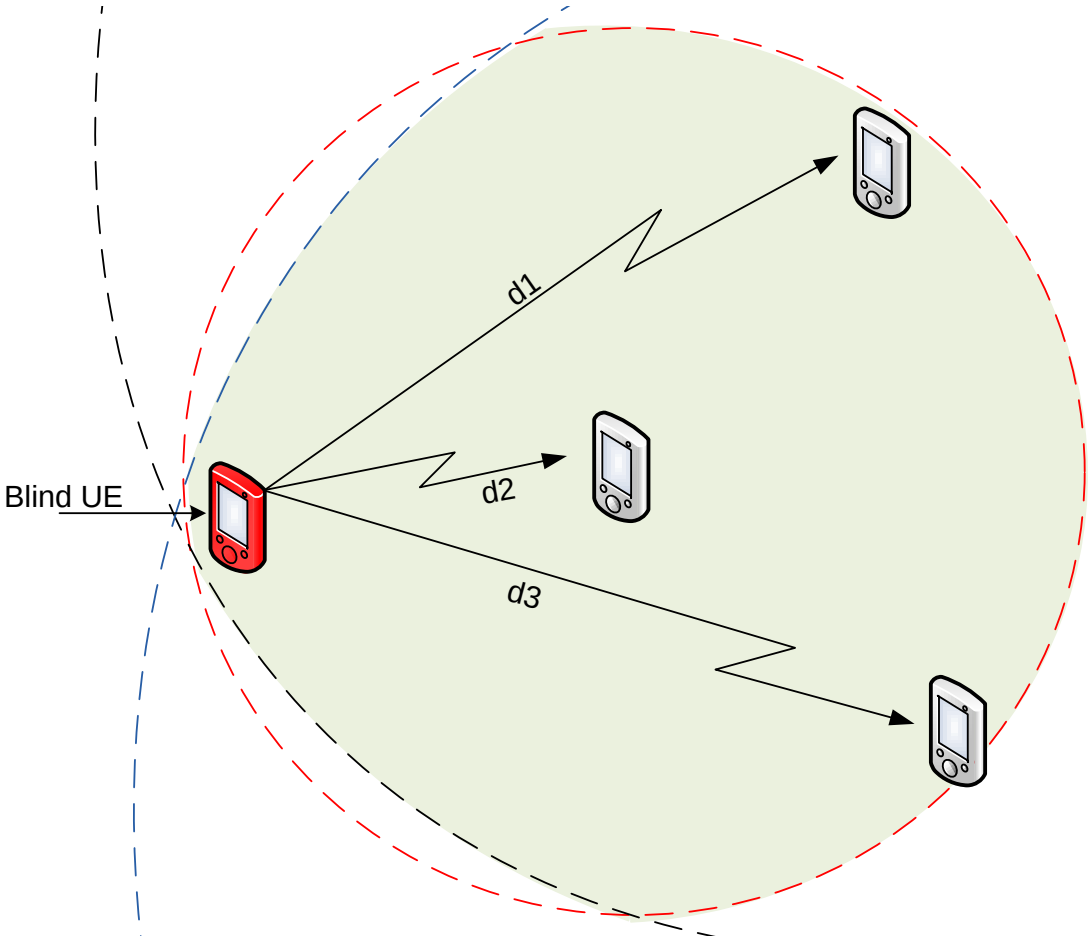


Figure 5.2: Measurement incertitude model in 2D.

We will accept incertitude of +/- e meters in order to solve our system of equations that will give as output a location zone for blind UE called measurement

incertitude represented in Fig. 5.2, where the white UEs are the already localized UEs and the red one represents the blind UE k.

If the system of equations provides no solution, i.e. no location zone, this means that one of the distance values does not match with the others, which may happen in case of strong error estimation of a distance. If this occurs the CH will look for another distance, called redundant distance, possibly available at the CH.

Corollary 5.2.1. *In the case of mapping failure in 2D, and in 3D due to error distance estimation ($n > 4$, $n > 6$ respectively), we benefit from additional $n - 4$, $n - 6$ resp., redundant distances to create a map.*

Proof. If we add UE n, we have $n - 1$ new distances, but we need from Proposition 5.2.2, only 3 distances, $n - 1 - 3 = n - 4$. □

The redundant distances are directly affected by the threshold P as more P is high, less redundant distances will be available. If no solution is provided by our system of equation, we reduce P. But even with such selection, the zone where blind UE k is possibly localized is still inaccurate. In the next section we will see these zones in detail and also how to make them more accurate.

5.3 Map Accuracy

In the previous section we ended up with a localization zone corresponding to a blind UE built by trilateration with errors due to ToA. The estimated blind UE position is normally the center of this zone, which is not accurate as shown in Section 5.2.2 Fig. 5.2. In this section, we will focus on improving the accuracy of the trilateration.

Algorithm 1 needs as input parameters the set of the three best distances $\mathcal{D}$ between the blind UE and the group $\mathcal{U}$ of already localized UEs. The goal of the algorithm is to tight this zone, where the first one is given by trilateration, by dividing it into sub zones and selecting the appropriated one. The procedure is based on dichotomy research where the value parameter for selection is represented by line 14 in Algorithm 1. After each iteration of this step we get as output a more accurate zone where its center is selected as the estimated location of the blind UE. All these steps are detailed in Algorithm 1.

Once applying Algorithm 1 to the previous measurement incertitude zone found in Section 5.2.2 Fig. 5.2, we see an improvement of the accuracy, represented by Fig. 5.3. In fact this example shows that with Algorithm 1 the estimation error was reduced from $d_{err} = 4.93$ m (first estimation given by the center of the fist zone) to $d'_{err} = 19.55$ cm (final estimation) showing an improvement of 96% accuracy. The channel model used for this example is indoor NLOS environment.

Algorithm 1 Map accuracy

Require: Estimated distances $\mathcal{D} = \{d_1, d_2, d_3\}$ between the new UE and the UEs already localized $\mathcal{U} = \{UE_i = (x_i, y_i) | i \in [1, |\mathcal{D}|]\}$
From $\mathcal{D}$ we get the estimated zone $\mathcal{Z}$ provided by the measurement incertitude in the general map
Compute center and perimeter of $\mathcal{Z}$: $c_{\mathcal{Z}}$ and $\mathcal{P}_{\mathcal{Z}}$ resp.
Estimated UE: $Est\_ue = c_{\mathcal{Z}}$
while $\mathcal{P}_{\mathcal{Z}} > \varepsilon$ **do**
  Divide the zone into 4 sub zones $\{\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3, \mathcal{Z}_4\}$ where border meet at $c_{\mathcal{Z}}$
  Compute their centers and perimeters $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$ and $\{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4\}$ resp.
  for $i = 1 : |\mathcal{C}|$ **do**
    $Sum_i = 0$
    for $j = 1 : |\mathcal{D}|$ **do**
      $d^2_{j,c_i} = (y_j - y_{c_i})^2 + (x_j - x_{c_i})^2$
      $Sum_i = Sum_i + (d^2_{j,c_i} - d^2_j)$
    end for
  end for
  $\mathcal{Z} = \mathcal{Z}_k$ and $\mathcal{P}_{\mathcal{Z}} = \mathcal{P}_k$ where $k = \arg\min_{k \in [1, |\mathcal{C}|]} Sum_k$
  $Est\_ue = c_z$
end while

In Section 5.1.1, we highlight one of the disadvantages of UWB which is its limited coverage range. In the next section we will see a possible enhancement for coverage extension.

5.3.0.1 UWB coverage extension and multi-group detection

As the coverage of UWB is limited up to around 20 m in outdoor environment, using frequency between 3 and 6 GHz [126], this range has to be increased in order to satisfy higher coverage specifications as for example for the use case that will be detailed in Section 5.5. One obvious option is by relaying the UWB signals. Let us now introduce UEs that have relaying capabilities. Our network is hence composed of four types of UEs:

- The CH, which is the one that launches the localization procedure and computes its corresponding map.
- The localized UE, which has been detected by a group. Its position is known by the CH. Once a localized UE belongs to a group, it stops listening to others group information signals.
- the blind UE, which has not been localized by any group and is still listening to group information signals.

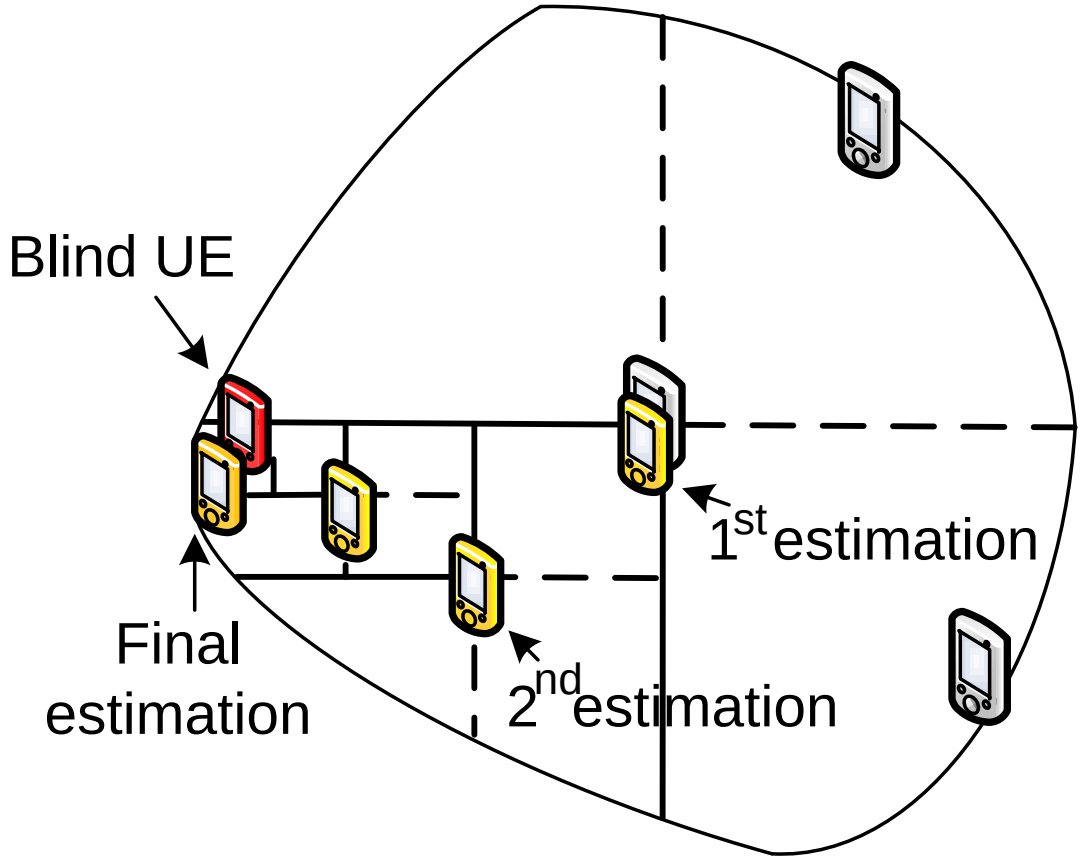


Figure 5.3: Blind UE positioning accuracy applying Algorithm 1.

- The relaying UE, that is localized at the border of its corresponding groups and is still listening to group information signals. Its goal is to relay the broadcast localization signal sent by the CH in order to extend the coverage. It is elected as relaying UE by any UE that is transmitting the group information broadcast signal, i.e. by a CH or any other relay.

We suppose that each UE is composed of a frame that is divided into three subframes, one for the group ID, one for the role of the UE (CH, relaying, localized or blind UE) and one for the number of hops to the CH.

As mentioned in Subsection 5.2, the CH is computing the localization map (see Fig. 5.1). Once the maximum UWB coverage rang is achieved and the map is performed, the CH will select the relaying UEs localized in the border of the group. All these UEs will be informed by the CH of their group ID, role (localized or relaying UE) and number of hops to the CH. Then, the relaying UEs will extend the

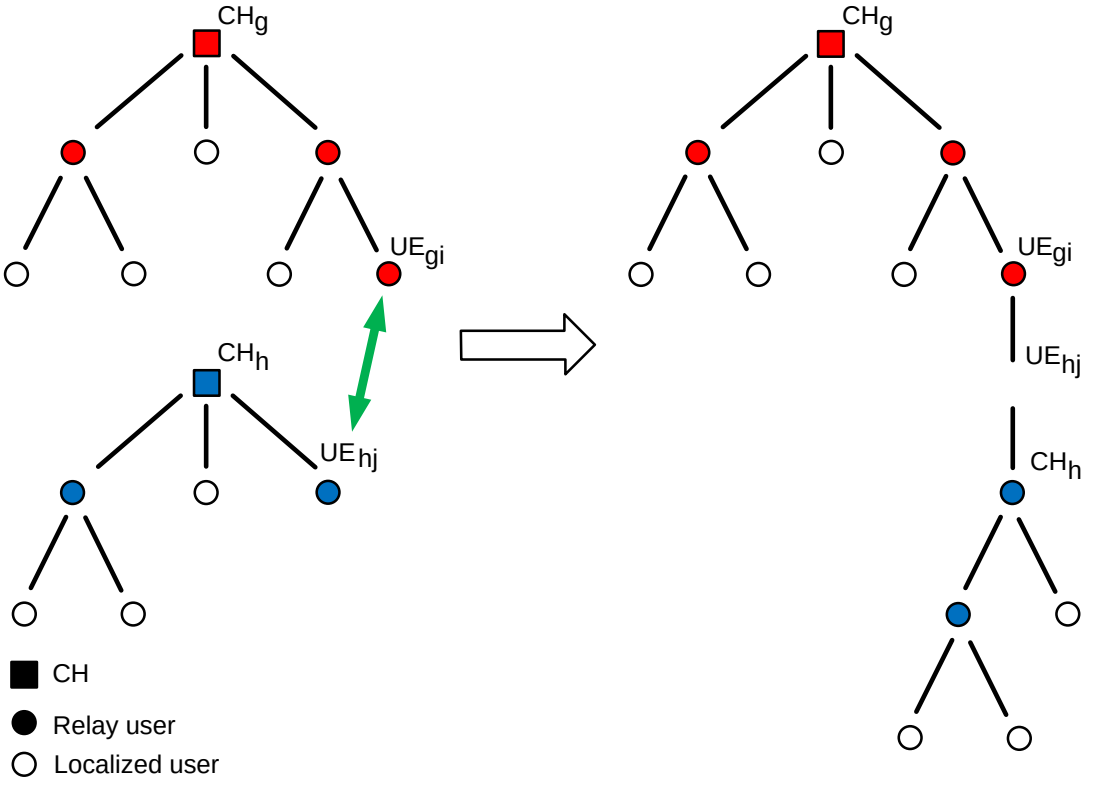


Figure 5.4: UWB coverage extension.

UWB coverage by retransmitting the group information broadcast signal intended to the still blind UEs, and repeat the same procedure until no UEs are localized.

In the case that a relay from group g detects a CH or a relay belonging to group h, this means that there may be one hotspot instead of two. In such a case, relay from group g and h will compare their corresponding hops subframe, and the one with less number of hops with its corresponding CH will change of group. Automatically, all the UEs attached to the relay that just made a group change will belong to the new group too. Their group ID subframe will change and the number of hops will be incremented. This is represented in Fig. 5.4, where UE_{gi} and UE_{hj} are two relays belonging to two different groups g and h respectively. One of these relays detects the other and they compare their subframe corresponding to the number of hops to their respective CH. Here the number of hops of UE_{gi} is 2 and the one of UE_{hj} is 1 so group h will be integrated into group g as represented in the figure. Also, the number of hops in the subframe of UE_{hj} is now 3 and the role of CH_h has changed into relaying UE and its corresponding number of hops is now 4.

In the next section we will see how accurate is our general method in four

different channel models: LOS/ NLOS for indoor and LOS/NLOS for outdoor.

5.4 Simulation Setup and Results

5.4.1 Simulation Setup

The models used bellow are respecting the IEEE 802.15.4a standards for UWB. In our case, as we are following the European regulation we selected the best bandwidth B which is 1 GHz centered in $f_c = 3993.6$ MHz where f_c represents the carrier frequency [127]. Let us recall the pathloss model defined in IEEE 802.15.4a standards for UWB:

$$PL = PL_0 + 10n_0 \log_{10}(d) \tag{5.1}$$

where PL represent the pathloss, PL_0 (dB) and n_0 are specific values depending on the studied model (LOS or NLOS for indoor or outdoor). d is the real distance between the transmitter and the receiver [126]. Recall that we use a bandwidth of $B = 1$ GHz. The SNR is expressed as

$$SNR_{dB} = P_t - PL - P_N \tag{5.2}$$

where P_t and P_N are the transmitted and noisy power respectively, and from the standards [126] their power density are equal to -41 dBm/MHz and -174 dBm/Hz respectively. Because we use estimated distances as input, we focused on the CRLB distance error which is one of the most commonly used measure of ToA accuracy [107]. From [121], the accuracy of ToA can be improved by increasing the SNR or the effective signal bandwidth. Since UWB signals have very large bandwidths, we benefit from extremely accurate location estimation using ToA via UWB radios. Also an approximation method of the CRLB when using UWB signals is introduced in [128] and another one using millimeter wave is proposed in [129]. The distance error is given by the CRLB d_{CRLB}, and the estimated distance is $d + d_{CRLB}$:

$$d_{CRLB} \approx \frac{c}{2\sqrt{2}\pi B\sqrt{SNR}} \tag{5.3}$$

For example, if we consider a distance $d = 10$ m to estimate in indoor LOS environment with the parameters above, we obtain a distance error $d_{CRLB} = 0.33$ cm. The measurements were done in four different IEEE 802.15.4a channel models: LOS and NLOS cases for indoors office environment and also outdoors [126]. Fig. 5.5 shows the estimation location of three blind UEs with four different channel models each. One has been zoomed in to clearly distinguish the location effect of the four channel models as well as the blind UE. In this case, the channel models with

strongest accuracy are LOS for indoor and outdoor and the one with poorest accuracy is NLOS outdoor as predictable. Simulation results for each channel model will be detailed in the next subsection.

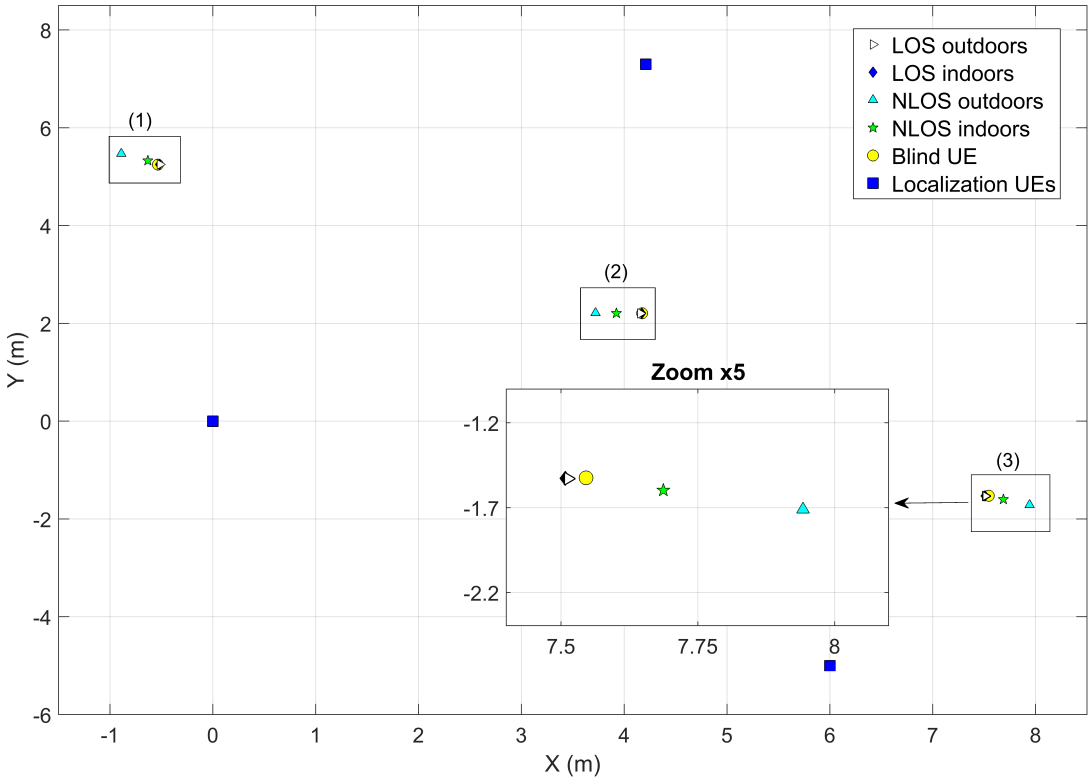


Figure 5.5: Three blind UE estimated localization for four channel models in 2D.

5.4.2 Simulation Results

In Fig. 5.6 is presented the repartition depending on the error between the blind UE position and the estimated position for the four different channel models. The simulation results show that the localization accuracy achieve its peak in LOS environments with an estimation error below 2 cm in 80% of the cases for LOS indoors and 73% for LOS outdoor. It is also important to note that half of errors estimation are below 1 cm in these two models. The accuracy is jeopardized in NLOS as the environments conditions are deteriorated, where 71% of the errors are below 20 cm in NLOS indoor and below 55 cm in NLOS outdoor.

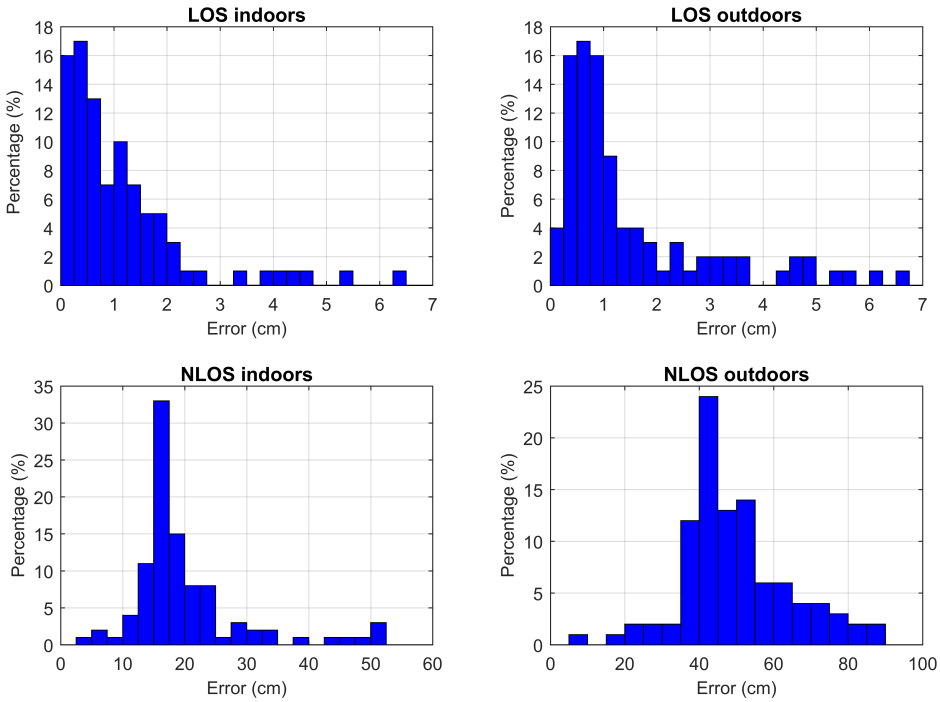


Figure 5.6: Algorithm simulations in four different UWB channel models.

5.4.3 Results Compared to other Methods

For indoors, we compare our results with other works who covered the same field. However, these works focused on resolving the most common problems of ToA signals [130, 131, 132, 133], meanwhile, we proposed a solution for the step after these ToA estimations. As we are working on two different visions we can only compare the final result which is the accuracy. The best accuracy we could find in the literature was of $3,9$ cm in perfect LOS [130] and less than 30 cm for NLOS indoors[131]. With the results presented in Section 5.4.2, we can conclude that we obtained more accurate results compared to other papers. We can add that these two visions can be combined to achieve even better results with our algorithm. For outdoors, we compare our results to GPS; as it remains the notorious outdoor localization technique and also because the literature lacks in the field of outdoors positioning for UWB. Our outdoors results have shown better accuracy (see Section 5.4.2) than GPS, whose accuracy varies from 2 m to 100 m [116]. Please note that an extension of GPS localization method is proposed in Section 5.5.1.1 in the context of cluster identification.

To summarize, we present an approach to create a map of group of smart devices by only using the distance between them, estimated by UWB ToA. Results show a large improvement of device localization with an average accuracy between 2 and 40 cm depending on the channel model. A direct application of this positioning algorithm can be used for cluster identification. The 4G considers the introduction of small cells in addition to the BSs to deal with dense traffic, as small cells help focusing energy on a given area such as commercial centers, stadiums, etc. Nevertheless, the use of small cells represents high investment for operators as they involves important CAPEX, i.e. backhaul deployment, site acquisition and also Operational Expenditure (OPEX), as energy consumption and maintenance [134, 135]. In order to avoid small cells deployment, 2D large antenna array at the BS named massive MIMO can be exploited. With the use of these 2D antenna arrays, highly directive beams in order to cover a given area can be created [136]. We call this concept VSC. UE positioning applied to 5G cellular management networks has been deeply described for different scenarios in [137] and also an architecture improving time response using location for indoor scenarios is presented in [138]. VSCs could be deployed in indoor scenarios, where nowadays small cells are more commonly used. The Home/office scenario is given by the constant and recurrent human behavior, i.e. switching from houses to offices during the day, to shopping malls and restaurant areas. Also, operators could cover specific floors in business buildings with VIP UEs having a punctual meeting or hosting an event. In this case, the traffic flow is predictive and therefore, operators can know in advance the occurrence of a high traffic demand in a given period of the day at a given location and program the activation and deactivation of VSCs pointing to the identified areas in the adequate periods. In crowded networks outdoor scenario, where human behavior is unpredictable a priori, identifying dense traffic areas in real time is more challenging and enables the network to dedicate them a specific VSC. Next section highlights the VSC concept and presents the benefits of a direct application of our positioning algorithm.

5.5 Group Localization aided Virtual Small Cells

The concept of VSCs is represented in Fig. 5.7, where is illustrated the deployment of the VSCs (on the bottom figure) that could replace the physical small cells (on the top figure), avoiding significant deployment cost for the operator. Adding more antennas at the BS provides more degree of freedom to the propagation channel between the transmitter and the receiver. Due to these degrees of freedom, higher diversity and data rate can be achieved, increasing also the capacity [139, 140, 141]. VSCs consist of highly directive beam working in co-channel that point to cluster estimated hotspot, increasing the SINR with massive MIMO at the BS and without

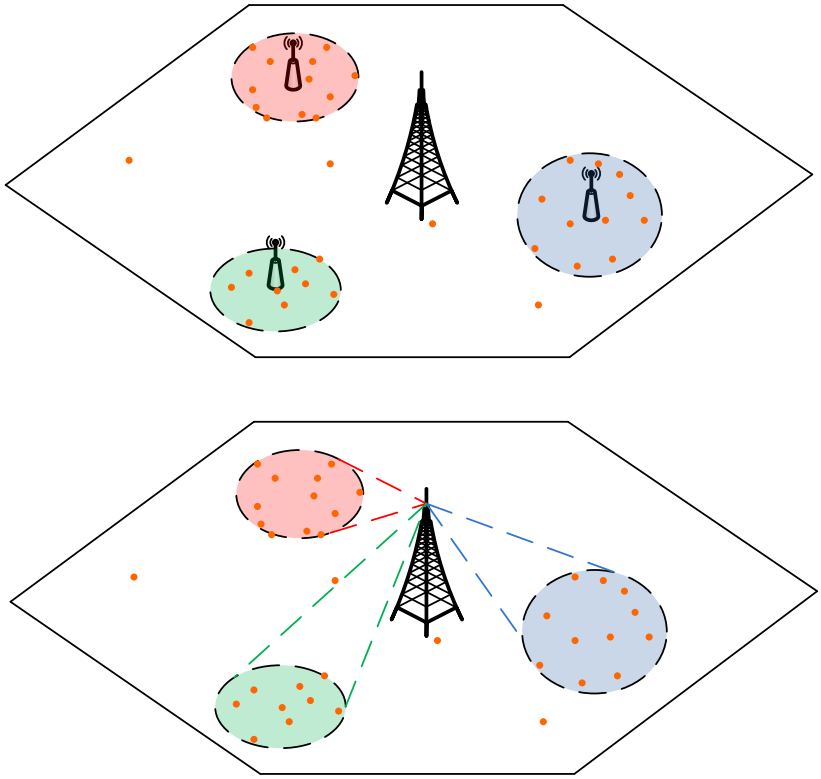


Figure 5.7: Network topology with and without VSCs example.

small cells [9]. Reduction of co-channel interference by using an unlicensed band for the VSCs is proposed in [142]. Security and caching aspects for VSCs have been considered in [143] and [144] respectively, while ergodic sum-rate has been derived in [145] for VSCs.

The benefits of VSCs compared to 3D UE-specific beamforming (for massive MIMO) are the following:

- VSC allows to dynamically allocate the radio resources for each geographical zone according to the dynamic spatial traffic.
- VSCs are transparent for UEs, and can then be used for legacy UEs. On the contrary, UE-specific beamforming requires UEs that support the number of extend codebooks.
- VSCs are more robust to UE mobility and/or channel variations since the focalization area is wider. If the 3D UE-specific beamforming is narrow, the motion of the UE forces to rapidly update the precoding in order to refocus the energy on the new UE location.

Note that there are two options when using a VSC. The first one is associating a different cell ID to the VSC with respect to the cell. This would correspond to network densification without new equipment deployment. The second one is to use the same cell ID (also same frequencies and mobility management) as that of the cell. In this case, the term VSC could be replaced by beam but would work similarly. This depends on the actual implementation of VSC, and particularly whether a different cell ID is given or not.

Regarding the implementation of the VSCs, the BS has to firstly evaluate the cellular traffic and secondly identify the hotspots where major part of the UEs are localized inside. If the detected hotspot is new in the cell, and it respects the constraint location, a highly directive beam will be focused on it, i.e. a VSC will be implemented. If hotspots in the database have disappeared, the associated VSCs are removed. By repeating these previous steps we obtain a periodical methodology which gives us a general view on when to implement or eliminate the VSCs. This methodology is summarized in Fig. 5.8.

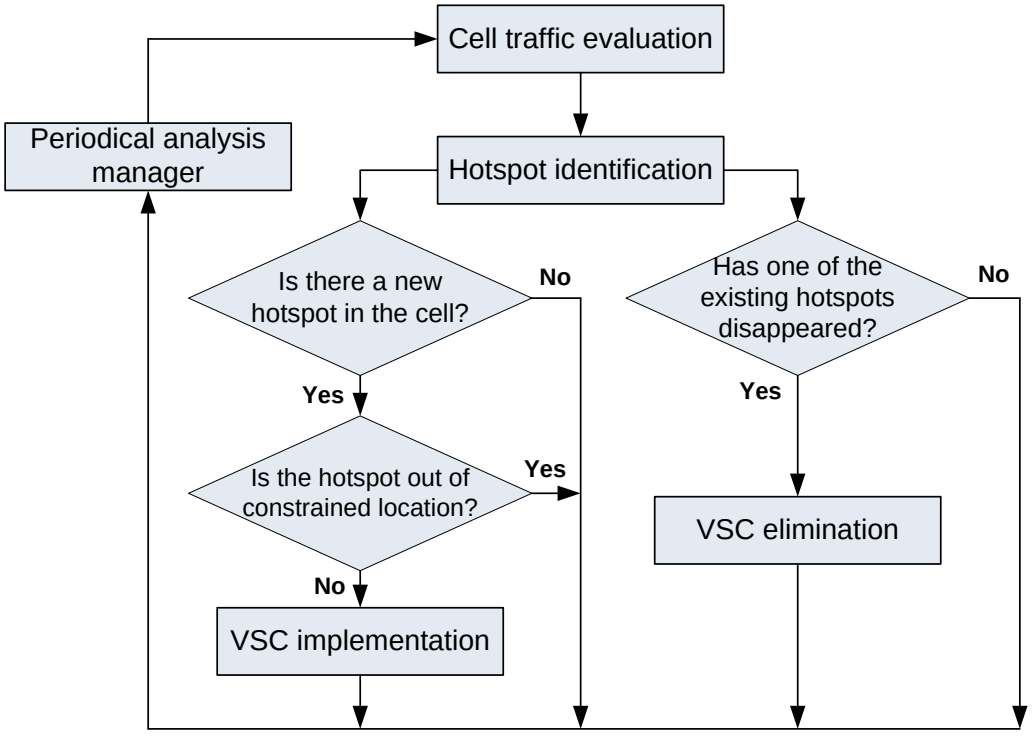


Figure 5.8: VSC configuration methodology.

5.5.1 Virtual Small Cell Architecture

This section introduces the relationship between group of UEs location and VSCs implementation. Considering that a hotspot has been identified, we will see that a common BS only requires angles information based on the localized hotspots in order to implement a VSCs, as long as the number of antennas is enough to steer the beam as required.

Fig. 5.9 describes the angles in the horizontal and vertical planes that are used for building a VSC. The VSC is hence defined by its steering θ_{tilt} and his beamwidth θ_{3dB} in the vertical plane as presented in the top Fig. 5.9. Those two angles can be respectively implemented using simple trigonometric functions:

$$\theta_{tilt} = -\arctan\left(\frac{h}{dist}\right) \ [rad] \tag{5.4}$$

$$\theta_{3dB} = \arctan\left(\frac{h}{dist - r}\right) - \arctan\left(\frac{h}{dist + r}\right) \ [rad] \tag{5.5}$$

where $dist$ is the projection in the vertical plane of the distance between the BS and the center of the localized hotspot, h is the height of the BS and r is the hotspot radius. From the horizontal plane we can distinguish two other angles, φ_{tilt} and φ_{3dB} as represented in the bottom Fig. 5.9 and expressed using trigonometric functions as follows:

$$\varphi_{tilt} = -\arctan\left(\frac{y - y'}{x - x'}\right) \ [rad] \tag{5.6}$$

$$\varphi_{3dB} = 2\arctan\left(\frac{r}{dist}\right) \ [rad] \tag{5.7}$$

where (x', y') are the coordinates of the located hotspot and (x, y) the coordinates of the BS.

The antenna elements (i.e. N_v and N_h antennas in the vertical and horizontal plane respectively), used for creating a beam, depends on the angles above. They also depend on the array element distance λD, where λ is the wavelength factor and on the value 50 that has been given by an antenna expert of our group based on experiments for our topology:

$$N_v = \frac{50}{\theta^o_{3dB}\lambda D\cos(\theta_{tilt})}, \tag{5.8}$$

$$N_h = \frac{50}{\varphi^o_{3dB}\lambda D\cos(\varphi_{tilt})}, \tag{5.9}$$

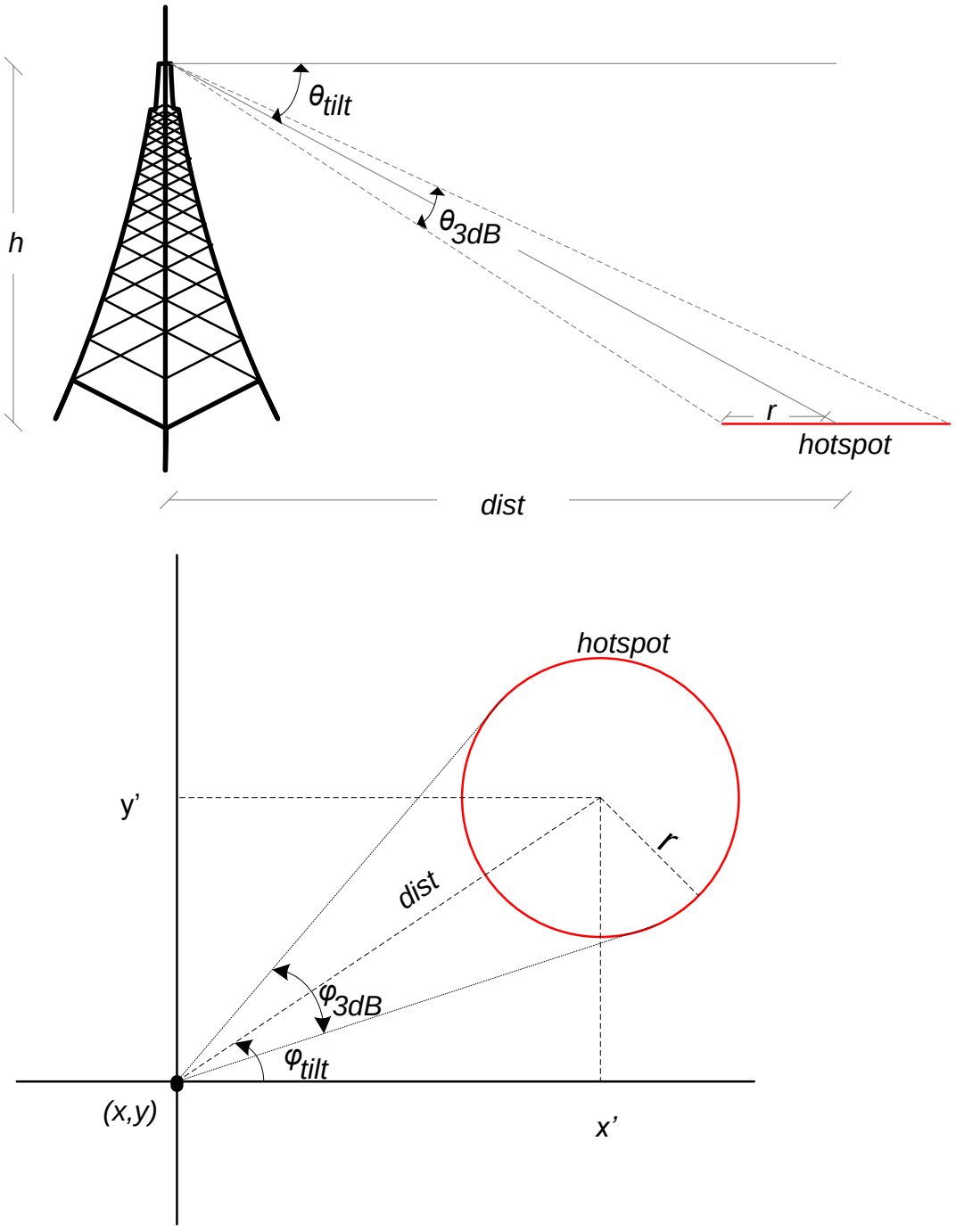


Figure 5.9: Beamwidth and beam steering in the vertical and horizontal planes.

Remark 5.5.1. *The performance of the VSC can vary depending on its steering point. There are basically two constraints to consider:*

- *If the VSC steering point is close to the BS, the UEs will experience high interference from the macrocell. θ_{tilt} will increase and its projection distance in the horizontal plane $d_{\theta_{tilt}}$ also, involving increased number of antenna elements in the vertical axis, as expressed in* (5.8)*.*
- *If the VSC is steering towards the edge of the cell, i.e. if the azimuth angle φ_{tilt} is close from bore-sight the number of antennas required is higher as shown in* (5.9)*.*

These two location constraints are represented in Fig. 5.10, where dashed areas represent the forbidden regions where VSCs will not be deployed.

Also from 3GPP Technical Report 36.814 [146], the antenna pattern is computed based on the 3D model as $A(\theta, \varphi) = -\min(-A_h(\varphi) + A_v(\theta), A_m)$, with maximum attenuation A_m, where:

$$A_v(\theta) = -\min\left(12\left(\frac{\theta_{tilt}}{\theta_{3dB}}\right)^2, A_m\right) \tag{5.10}$$

for the vertical antenna pattern, and for the horizontal one:

$$A_h(\varphi) = -\min\left(12\left(\frac{\varphi_{tilt}}{\varphi_{3dB}}\right)^2, A_m\right) \tag{5.11}$$

and the pathloss models used for NLOS is the one defined in the 2D 3GPP model:

$$PL = 131.1 + 42.8\log_{10}(R) \tag{5.12}$$

where $R^2 = dist^2 + h^2$ in km from Fig. 5.9.

Remark 5.5.2. *It is important to note that all the information needed by the BS in order to implement the VSC only depends on angles metric represented in equations* (5.4) *to* (5.7) *and on distance values. These metrics can be deduced by knowing only the coordinates and the radius of the desired hotspot, as height and coordinates of the BS are already known. Hence easy precoding can be implemented, where a predefined codebook will be based on specific characteristics of the detected hotspot. These characteristics are the center and the radius of the corresponding hotspot.*

Remark 5.5.3. *A VSC system model from radio point of view also depends only on the mentioned angles information, such that the proposed precoding is still applicable. For more details regarding the VSC's radio model and associated codebook please refer to [147].*

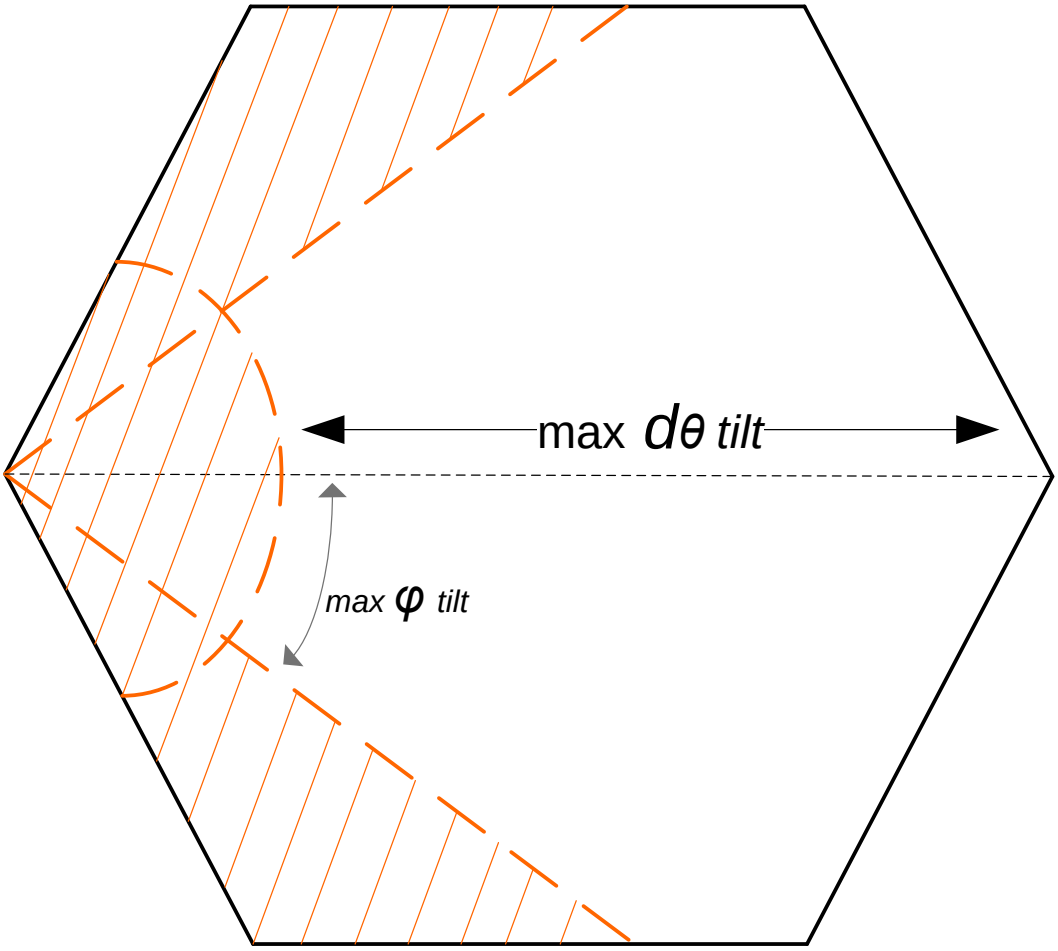


Figure 5.10: Constraint location for VSC in hexagonal cell.

As described in Remark 5.5.2, to be able to implement the VSCs, we have to characterize their center and their radius. Depending on operator strategies, several clustering methods have been developed for positioning in 5G [148] in order to identify such a hotspot. One strategy is to identify the hotspots without intervention of the BS by using direct sidelinks between all the UEs such that the BS is not overcharged by the UEs data, as presented at the beginning this chapter. Numerical results when applying our D2D aided group localization method to the hotspot identification for VSCs will be presented in Section 5.5.1.2, where a comparison with another strategy is proposed. This second strategy is a centralized positioning algorithm at the BS as presented in next section.

5.5.1.1 UEs Mapping using Global Positioning System

Here we respond to the missing point of hotspot location by proposing a dynamic clustering algorithm based on GPS coordinate location. GPS is the most commonly used localization technique, in particular for outdoor scenarios. We suppose that the BS benefits from Minimization of Drive Tests (MDT) [149] which collects network quality information like UE coordinates, i.e. MDT geo-localize the UEs. Detailed location information provided by MDT reports, allows the operator to associate a set of MDT measurements with a physical location. The UEs are requested by the BS to acquire location information for a configured MDT session [150]. The knowledge of the UEs coordinates is hence available at the BS and can be used in order to find the hotspots. The BS will apply an algorithm based on the well-known K-means that outputs or create clusters of UEs. Then, after some optimizations we succeed on transforming the clusters into hotspots. The general algorithm for hotspot detection is constructed as in [115]. First, K-means is an unsupervised learning algorithm that solves clustering problem. In K-means, a set of K data points (replaced by n in the algorithm, as it represents the number of UEs) and an integer C which represents the number of centers to evaluate are given. These centers are randomly defined and then optimized in order to minimize the mean squared distance from each data point to its nearest center [151], as described between the lines 1 and 15 in Algorithm 2. Then, the optimal number of clusters is selected based on the distortion function provided by [152], as explained in Algorithm 2 in lines 16 to 23. Secondly, to avoid UEs located far from the cluster and still grouped to it by K-means, a process that removes the far UEs is implemented, which is presented between lines 24 to 32. Also, please notice that if a cluster is considered as low populated it will be removed from the database, where p_1 is a parameter configurable by the operator. The last optimization is on determining if the cluster can be smaller covering almost the same number of UEs by looking each cluster density thanks to their distortion values. Parameter p_2 line 37 represents a fraction of the Inter Site Distance (ISD).

In order to evaluate the potential of our algorithm, let us consider a system composed of 57 hexagonal cells with ISD of 500 m. Then in each cell we have "created" a circle of 40 m radius and randomly placed 60% of the UEs inside and 40% outside this circle, as represented in Fig. 5.11. We consider perfect GPS localization. The error GPS localization will be taken into account in Section 5.5.1.2. When applying the proposed Hotspot Identification algorithm to each cell and by repeating the process 100 times, we obtain a probability of 55% of detecting one hotspot per sector, 41% of detecting two hotspots and 4% of detecting three hotspots. This means that new groups of UEs have been detected. Also, the Hotspot Identification algorithm is covering 72% of the total population which is

Algorithm 2 Hotspot Identification

BS requests for immediate MDT data from the subscriber UE
Input: n, $(x_k, y_k)_{k=\{1,\dots,n\}}$, $C_{max} = 3$
for $C = 1 : C_{max}$ **do**
  $l = 0$
  $\mathbf{M}_C^{l-1} = 0$, $\mathbf{M}_C^l = \begin{bmatrix} \mathbf{m}_1^l \\ \dots \\ \mathbf{m}_C^l \end{bmatrix} \in \mathbb{R}^{C\times 2}$ with $\mathbf{m}_1^l$ random center
  while $\mathbf{M}_C^l \neq \mathbf{M}_C^{l-1}$ **do**
    $\mathbf{M}_C^{l-1} = \mathbf{M}_C^l, \mathcal{U} = \emptyset, l = l + 1$
    Determine nearest center for each UE
    for $k = 1 : n$ **do**
      $j = \arg\min_{i=1,\dots,C} \sqrt{(y_k - m_{i,2}^l)^2 + (x_k - m_{i,1}^l)^2}$
      $\mathcal{U}_j = \mathcal{U}_j \cup \{(x_k, y_k)\}$
    end for
    for $j = 1 : C$ **do**
      $\mathbf{m}_j^l = \sum_{k\in\mathcal{U}_j} \frac{(x_k,y_k)}{|\mathcal{U}_j|}$, $\mathbf{M}_C^l = \begin{bmatrix} \mathbf{m}_1^l \\ \dots \\ \mathbf{m}_C^l \end{bmatrix}$
    end for
  end while
  Estimate the distortion for each cluster:
  for $i = 1 : C$ **do**
    $I_i = \sum_{k\in\mathcal{U}_i} \sqrt{(y_k - m_{i,2}^l)^2 + (x_k - m_{i,1}^l)^2}$
  end for
  Total distortions: $S_C = \sum_{i=1}^{C} I_i$
  Determine estimation function $f(C)$ with d (dimension) and α_C (weight factor)

$$f(C) = \begin{cases} 1 & \text{if} \quad C = 1 \\ \frac{S_C}{\alpha_C S_{C-1}} & \text{if} \quad S_{C-1} \neq 0, \forall n > 1 \\ 1 & \text{if} \quad S_{C-1} = 0, \forall C > 1 \end{cases} \tag{5.13}$$

$$\alpha_C = \begin{cases} 1 - \frac{3}{4d} & \text{if} \quad C = 2 \text{ and } d > 1 \\ \alpha_{C-1} + \frac{1+\alpha_{C-1}}{6} & \text{if} \quad C > 2 \text{ and } d > 1 \end{cases} \tag{5.14}$$

end for
Select the optimal number of clusters:
$C_{opt} = \arg\min_{C_i=1,\dots,C_{max}} f(C_i)$, and let $\mathbf{M}_{C_{opt}}$
for $C = 1 : C_{opt}$ **do**
  Compute average distance $\mu_C = \frac{1}{|\mathcal{U}_C|} \sum_{k\in\mathcal{U}_C} d_{k,C}$
  where $d_{k,C}$ is UE k distance from his center C
  Compute standard deviation:
  $\sigma_C = \sqrt{\frac{1}{|\mathcal{U}_C|} \sum_{k\in\mathcal{U}_C} (d_{k,C} - \mu_C)^2}$
  for $k = 1 : |\mathcal{U}_C|$ **do**
    if $d_{k,C} > (\mu_C + \sigma_C)$ **then**
      Remove UE k from $\mathcal{U}_C$
    end if
  end for
  if $|\mathcal{U}_C| < p_1 n$ where $p_1 = 15\%$ **then**
    remove cluster C
  end if
  Compute the average distortion of each cluster $\bar{I}_C = \frac{I_C}{|\mathcal{U}_C|}$
  if $\sqrt{\bar{I}_C} > p_2$ **then**
    Recompute this Algorithm from line 25
  else
    Compute cluster radius $\mathbf{r}_{C_{opt}}(C) = r_C$
  end if
end for
Output: $\mathbf{M}_{C_{opt}}$ and $\mathbf{r}_{C_{opt}}$

12% more than the considered cellular system. Furthermore, after repeating the process hundred times, half of the radius measured are between 30 and 45 m, as it can be seen in Fig. 5.12.

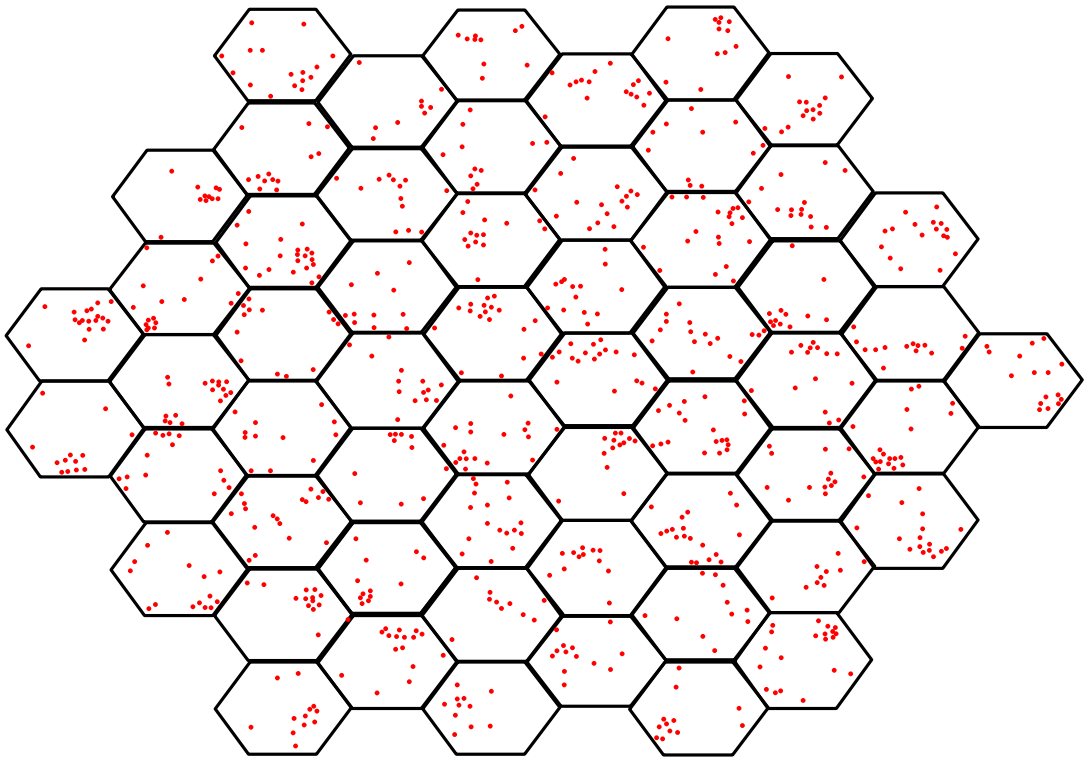


Figure 5.11: UEs distribution in the network.

In Fig. 5.13, 25% of the radius are below $32,7$ m, half of the radius are below $39,4$ m and 75% of the radius are below 50 m. A nearly Gaussian distribution for the cluster radius is observed which means that the closest values from the mean are more able to appear than the others. This is satisfactory as half of the clusters radius are close to 40 m.

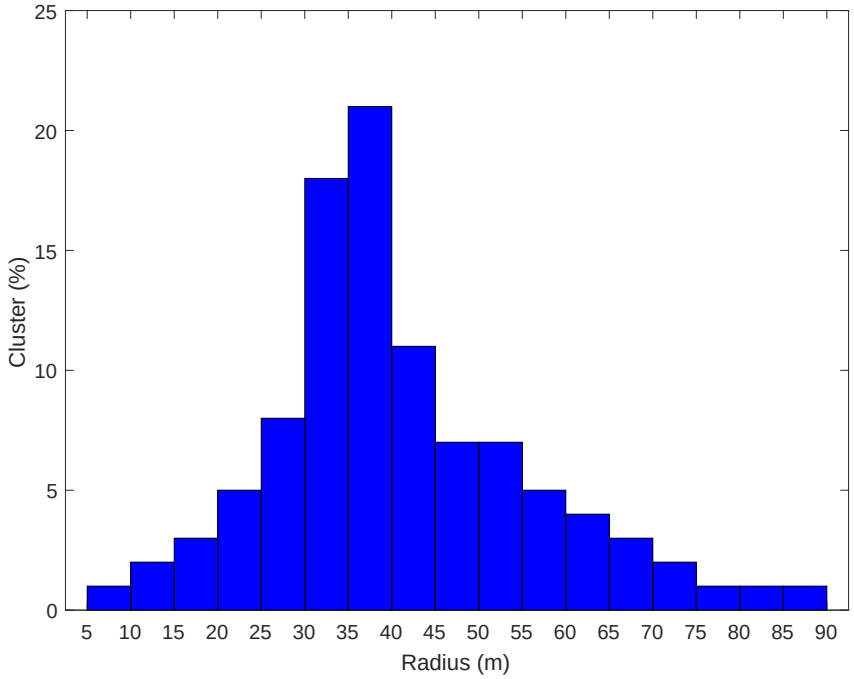


Figure 5.12: Cluster radius distribution.

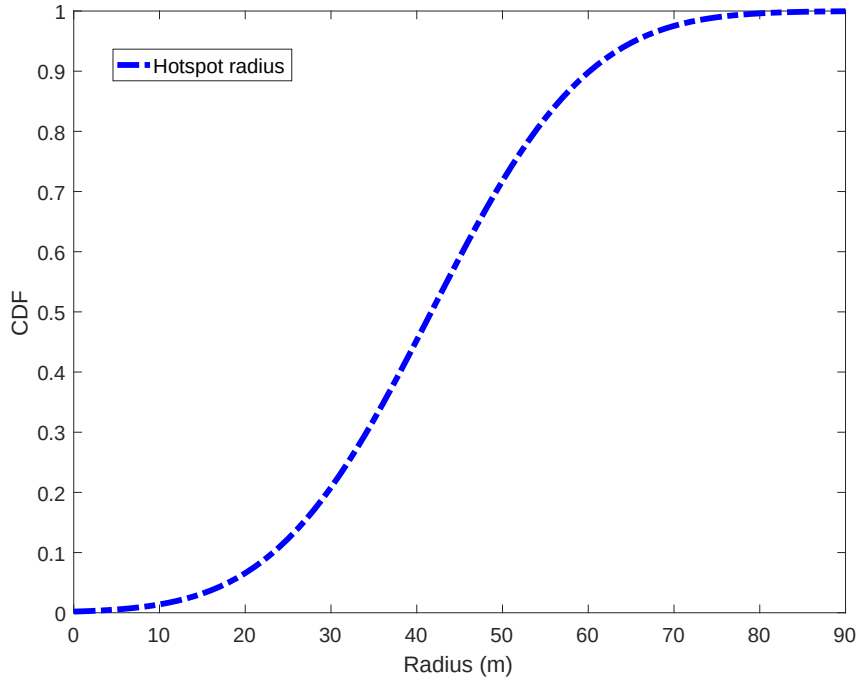


Figure 5.13: Statistics on hotspots radius.

Now that the centralized positioning algorithm at the BS using GPS has been presented, we can compare its results to the ones provided by the D2D aided group localization, where part of Algorithm 2 has been applied in order to transform the map of a group of UEs into a map of hotspots.

5.5.1.2 Comparison

As our hotspot localization methods are intended for the use of the VSCs, we consider a cellular system with 150 UEs distributed in such a way that in each sector 60% of the UEs are clustered in a 40 m hotspot radius and the rest are randomly distributed (i.e. like if there was a pico-cell in a cellular sector). We apply our two hotspot localization methods and highlight four representative clustering realizations depending on UEs distribution, as shown in Fig. 5.14.

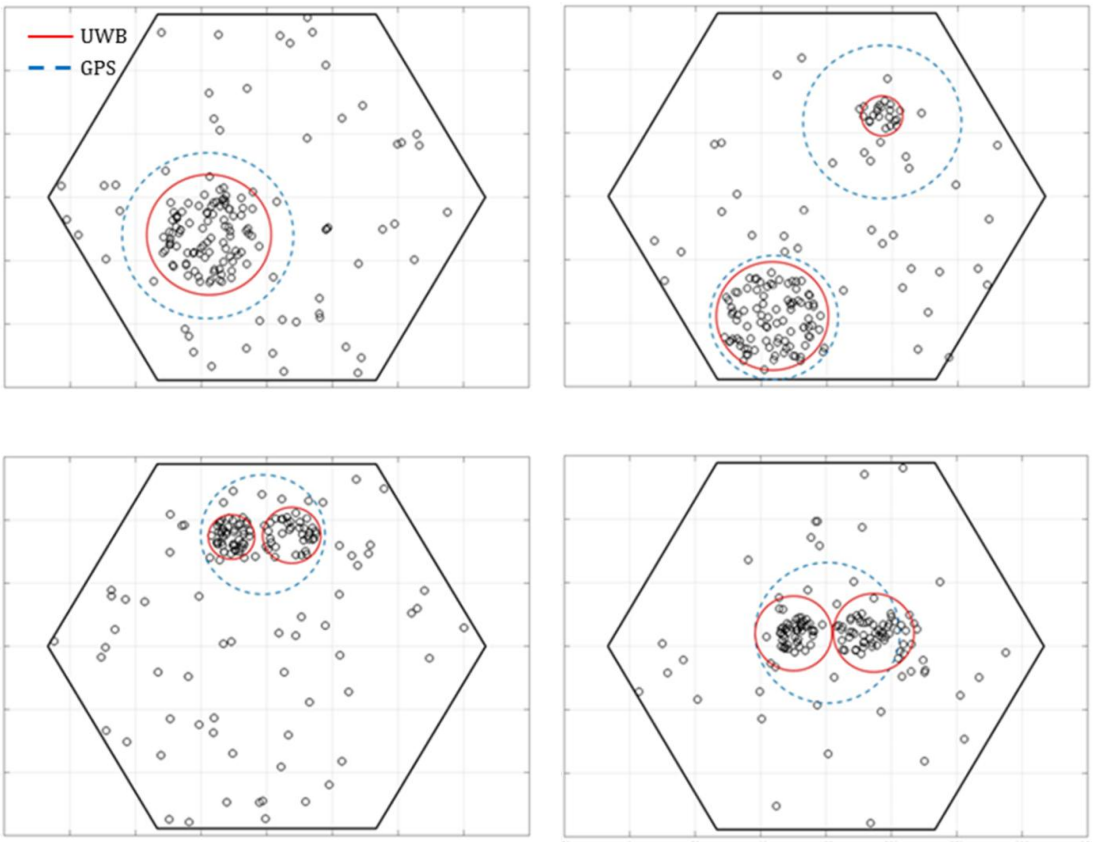


Figure 5.14: Four representative clustering realizations.

In realization i, presented in the top left Fig. 5.14, both methods give us a hotspot where the one given by UWB signals has a smaller radius than the one given by the GPS, in particular due to the GPS error localization of 10 m in average [153]. Top right Fig. 5.14 highlight a second representative realization where the 40% of the UEs, which are outside the predefined hotspot, are distributed in such a way that they form a second hotspot in the cellular sector. Meanwhile both clustering methods detect this hotspot, its radius is much smaller using UWB method than GPS and also seems to be more dense. Bottom left and right figures (in Fig 5.14) illustrate a quite similar repartition as in both cases the 60% of the UEs forming the predefined hotspot of 40 m radius can actually be separated into two sub-hotspots. GPS hotspot detection is able to detect the predefined hotspot only, whereas UWB detects the two sub-hotspots. In realization iv, UEs close to the predefined hotspot are taken into account by the UWB method, and are included into the final detection.

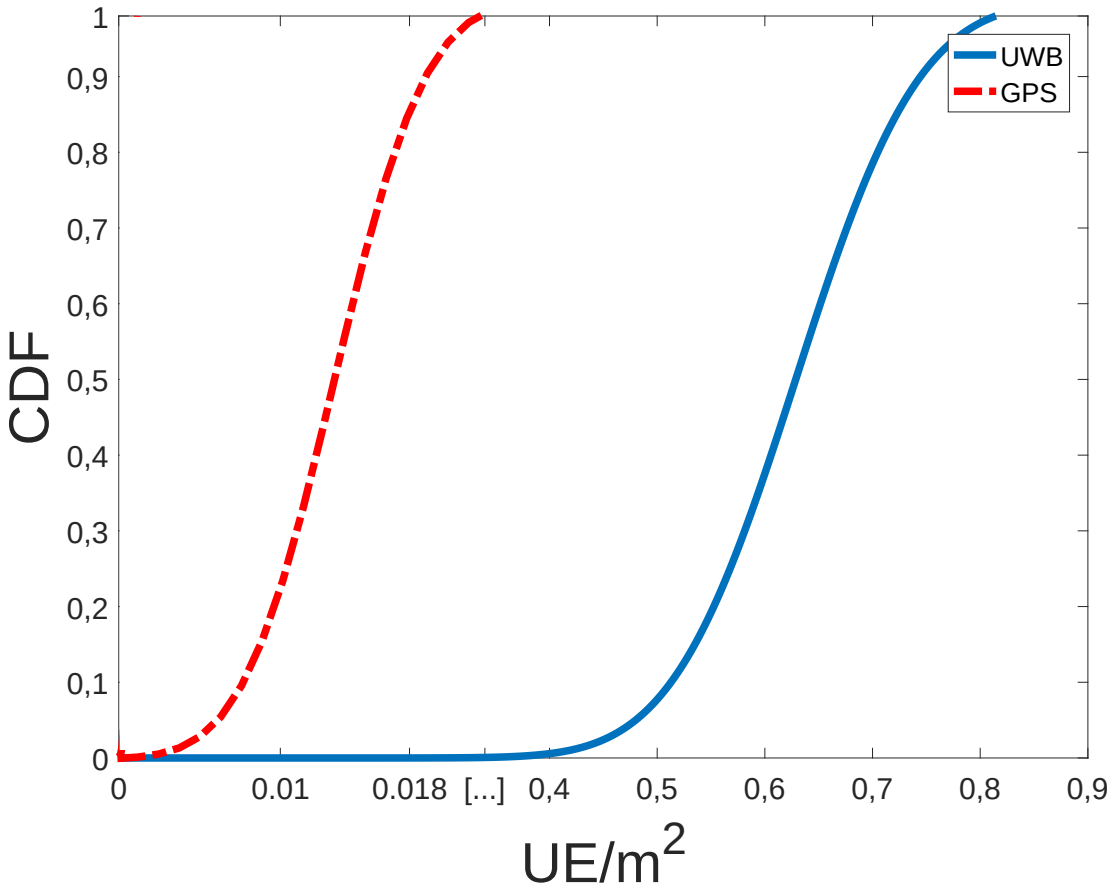


Figure 5.15: CDF of the UE density per m^2 depending on the localization method.

Fig. 5.15 represents the Cumulative Distribution Function (CDF) of the UE density per square meter for the detected hotspots with both methods and we observe that UWB provide much more dense hotspots than GPS method.

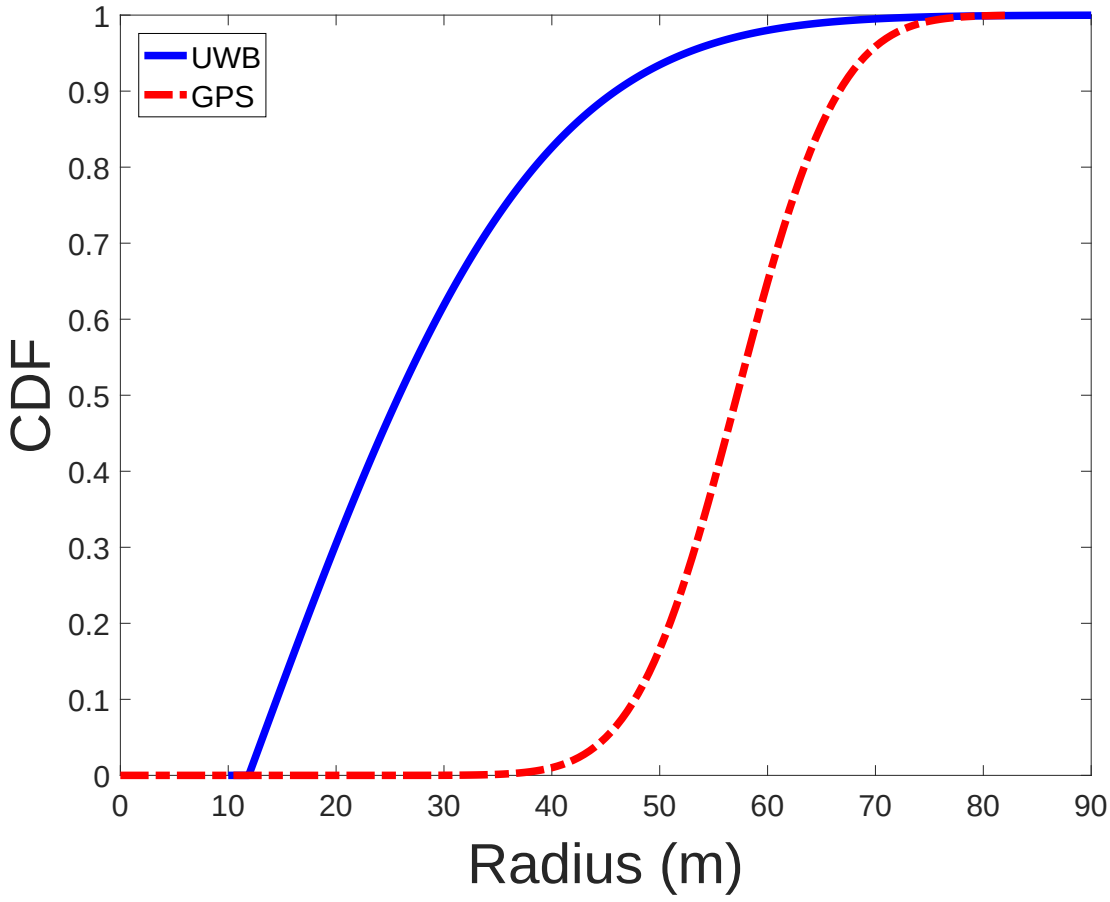


Figure 5.16: CDF of the hotspots radius depending on the localization method.

Also, it is important to note that the hotspots radius are in general smaller using UWB than using GPS as it can be seen in Fig. 5.16, where the CDF of the hotspots radius is represented for both methods. Generalizing, the more precise the localization system is, the better the clustering results are. Nevertheless, when selecting localization method other criteria should also be taken into account such as the complexity, battery consumption, terminal cost, etc. As our objective is to implement VSCs based on hotspot localization, it is important to focus the beamforming signal on a dense and accurate hotspot such that the energy and the resources used while performing the VSCs are optimal.

5.6 Conclusion

In this chapter, we present an approach to create a map of group of smart devices by only using the distance between them, estimated by unlicensed UWB signals in D2D outband and ToA ranging technique. The devices are using UWB signals as it remains the more accurate and less affected by obstacles due to its large band in the high frequency domain. Simulations applying this method show a large improvement of the device localization for a 1 GHz bandwidth, with an average accuracy between 2 and 40 cm depending on the channel model. This chapter presents also the concept of VSC as an alternative to classical heterogeneous network deployed in 4G, 5G and beyond, avoiding the deployment of small cells that implies a non-negligible cost in terms of new equipment deployment, sites acquisition and maintenance for operators. Their advantage is that they are reconfigurable during the time and are flexible to changing traffic conditions. A second method has been presented in order to dynamically localize the hotspots in the network that is centralized at the BS by using UEs GPS coordinates. This method shows to correctly detect the optimal number of hotspots at a given time. Both localization methods have been compared in an outdoor LOS environment and the D2D based method provides better accuracy results, with error estimation smaller than 1 cm in 50% of the cases. The GPS based localization method can still be used for VSCs in case where no direct communication between the devices is available.

5.7 Appendix

We recall below the list of parameters used along the chapter:

Table 5.1: List of parameters in Chapter 5

n	number of UE
(x_i, y_i)	2D coordinates of UE i
(x_i, y_i, z_i)	3D coordinates of UE i
(x, y)	2D coordinates of BS
(x', y')	2D coordinates of hotspot center
$\mathcal{U}$	set of localized UEs
C_{max}	maximum number of possible hotspots
h	height of the BS
r	hotspot radius
$d_{i,j}$	distance between UE i and UE j
d_{CRLB}	distance error given by CRLB
N_v	number of antennas in the vertical plane
N_h	number of antennas in the horizontal plane
λ	wavelength factor
D	distance between 2 antenna elements
θ_{tilt}	beam steering angle in the vertical plane
$\theta_{3_{dB}}$	beamwidth in the vertical plane
φ_{tilt}	beam steering angle in the horizontal plane
$\varphi_{3_{dB}}$	beamwidth in the horizontal plane
$d_{\theta_{tilt}}$	projection distance of θ_{tilt} in the horizontal plane
A_m	maximum antenna attenuation
$A_v(\varphi)$	vertical antenna pattern
$A_h(\varphi)$	horizontal antenna pattern
P_k	received power by UE k
P	received power threshold
P_t	transmitted power
P_N	noisy power
f_c	carrier frequency
B	bandwidth
PL	pathloss

Chapter 6

Conclusion and Perspectives

In this thesis, we explore how Device-to-Device (D2D) technology enhances the performances in three critical use cases for Fifth-Generation (5G). These scenarios highlight how D2D communication is used for enabling the three 5G major classes of services: enhanced Mobile BroadBand (eMBB), Ultra-Reliable and Low Latency communication (URLLC) and massive Internet of Things (IoTs).

Within the eMBB class of services (e.g. video streaming), we study the multicast wireless scenario in Chapter 3. We studied a new scheme for multicasting common message to User Equipments (UEs) based on D2D communication between the receivers and without Channel State Information at the Transmitter (CSIT). We have derived a complete characterization of the proposed scheme that guarantees a multicast rate of $\mathcal{O}(\ln \ln n)$ in a regime of large number n of UEs. For finite systems, we also have provided tractable and accurate approximations even for systems of modest sizes. We have also proposed two extensions that increase the performance of our scheme: (i) by optimizing resource allocation and (ii) by considering the use of Hybrid Automatic Repeat Request (HARQ).

The second wireless scenario studied in Chapter 4 of this thesis is D2D communication under reliability and latency constraints. This scenario is part of the second major class of services (URLLC) such as Vehicular-to-Everything (V2X) or industry automation. Here we studied the problem of designing HARQ schemes in systems which have both latency and reliability constraints. In the case of a single UE transmitter in a noisy environment, we derive the optimal policy and prove that our semi-greedy policy (i.e. which corresponds to transmitting one packet every slot interval of delay feedback and then transmitting the remaining packets as a burst in the last slots) is a good choice in practice, and has strong theoretical performance guarantees, in some regimes of interest. Furthermore, we have proposed policies with near-optimal performance for multiple UE transmitters scenario, as well as a mean field analysis to predict their performance.

Finally, the last scenario explored in Chapter 5 of this thesis is the one of localization in the context of massive IoT network like smart cities or industry 4.0. In this context, we present an approach to create a map of group of smart devices by only using the distance between them. We also present the concept of Virtual Small Cell (VSC) aided by hotspots location prediction in real time, as an alternative to the deployment of new small cells infrastructure, implying non-negligible cost in terms of new equipment, sites acquisition and maintenance for operators.

Despite the fact that these three D2D based schemes appear as enabler for the three major classes of 5G services (eMBB, URLLC and massive IoT), there are still several challenges to investigate in the future.

First of all, the schemes presented in this thesis represent realistic scenarios. Therefore, if we want all D2D UEs of the network to be able to participate in such schemes, they need to be specified in 3GPP standards.

Regarding the multicast scheme of Chapter 3, the two-phase transmission procedure (i.e. the BS transmits at a given rate in the first phase, and remains silent during the second phase while the D2D UEs perform the retransmission) can contribute to 3GPP standardization. Indeed, all the operators will benefit from this increasing multicast rate and it will be especially interesting for Public Safety (PS) UEs like firefighters. Furthermore, the inputs needed by the operators in such a protocol represent small amount of data. In addition, our scheme has several other advantages, it is distributed, with little feedback (only statistical knowledge), it is scalable with the densification of the network and it guarantees strong performances compared with the common multicast scheme (whatever the number of UEs present in the network).

For the work related to URLLC, presented in Chapter 4, as 3GPP already adopts the multiple replicas transmission without waiting for the Acknowledgments (ACKs). One could propose our retransmission policy, in the case of single transmitter, in the continuity of their standardization. Our retransmission policy does not depend on the loss rate of the channel, it is easy to implement and it can be adapted to random feedback delays (depending on the service). This will increase the reliability of the already existing schemes, especially in the new context of V2X or industry automation. There are already study items and work items dealing with V2X and industry automation which could benefit from our proposal.

Finally, regarding the last chapter, one could think about standardizing the localization method and study how to integrate it into 5G network. Notice that the tendency of increasing the bandwidth all along the upcoming network generations matches our hypothesis of ultra large bandwidth. Discovery signals could be reused to measure the distances. The algorithm to find the location using such

distances does not need to be specified, and each vendor could implement it in a different way. What remains to be specified is the broadcast messages described in Chapter 5, with a standard format and size so that they are understood by every D2D UE.

The further improvements of the thesis' work are proposed in the following.

Concerning D2D aided multicast scenario (see Chapter 3), one could formally express the upper bound of our proposed scaling multicast rate from an information theoretic point of view in order to answer the question: is our scheme optimal? Moreover, a possible future work is to extend the study for multiple antennas at the transmitter and analyze the impact of D2D communication. Finally, several services like content delivery are strongly affected by the common vanishing multicast rate when the number of UEs is increasing. Therefore, it could be of great interest to adapt our increasing D2D aided multicasting rate for such services (i.e. by providing a scalable content delivery based on coded caching).

Another future work for D2D aided URLLC scenario (see Chapter 4), could be to find the optimal distributed policy in the multiple UE transmitter case, and to prove that our mean field analysis can be made rigorous, since it does predict the system's behavior accurately.

One could think of combining these two schemes, by considering a single UE that wish to share a common message to the whole group by multicast. Still considering the two-phase scheme of Chapter 3, what could be the optimal retransmission policy under URLLC constraints?

Regarding now D2D aided localization (see Chapter 5), we could reduce the number of measurements in practice by introducing the relative Angle of Arrival (AoA) in our proposed method. Another possibility could be to study some algorithm of group movement prediction as a way to enhance network coverage.

Chapter 7

Appendix

7.1 List of Publications

7.1.1 Journal Papers

1) [154] T. Varela Santana, R. Combes and M. Kobayashi, "Device-to-Device Aided Multicasting", *submitted to* IEEE Transactions on Wireless Communications, 2018.

2) [147] T. Varela Santana, S. Martinez Lopez and A. Galindo-Serrano, "The Virtual Small Cells based on UE Positioning: A network densification solution", EURASIP Journal on Advances in Signal Processing, 2018.

7.1.2 Conference Papers

3) [155] T. Varela Santana, R. Combes and S.E. El Ayoubi, "On ARQ with Delayed Feedback under Latency and Reliability Constraints", *submitted to* IEEE INFOCOM, Paris, 2019.

4) [156] T. Varela Santana, R. Combes and M. Kobayashi, "Device-to-Device Aided Multicasting", IEEE International Symposium on Information Theory (ISIT), Vail, 2018, pp 771-775.

5) [157] T. Varela Santana, S. Arrefag and S. Martinez Lopez, "A High Resolution Method for Equipment Group Mapping Using UWB Signals", IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Montreal, 2017, pp 1-5.

6) [115] T. Varela Santana, A. Galindo-Serrano, B. Sayrac and S. Martinez Lopez, "Dynamic network configuration: Hotspot identification for Virtual

Small Cells", IEEE International Symposium on Wireless Communication Systems (ISWCS), Poznan, 2016, pp 49-53.

7.1.3 Patent

7) [158] T. Varela Santana and S. Martinez Lopez, "Procédés et dispositifs de cartographie flottante d'un groupe de mobiles", patent n^o FR3065142, published 12/10/2018.

7.2 List of Figures

7.3 List of Tables

7.4 List of Acronyms

1G First-Generation

2D Two-dimension

3D Three-dimension

2G Second-Generation

3G Third-Generation

3GPP 3rd Generation Partnership Project

4G Fourth-Generation

5G Fifth-Generation

ACK Acknowledgment

AoA Angle of Arrival

AWGN Additive White Gaussian Noise

BS Base Station

CAPEX Capital Expenditure

CC Chase Combining

CDF Cumulative Distribution Function

CDMA Code Division Multiple Access

CH Cluster Head

CRLB Cramer-Rao Lower Bound

CSIT Channel State Information at the Transmitter

CSMA Carrier Sense Multiple Access

D2D Device-to-Device

eMBB enhanced Mobile BroadBand

eMBMS enhanced Multimedia Broadcast Multicast Service

FDD Frequency Division Duplexing

GPS Global Positioning System

GSM Global System for Mobile communications

HARQ Hybrid Automatic Repeat Request

IEEE Institute of Electrical and Electronics Engineers

i.i.d. independent and identically distributed

IoT Internet of Thing

IR Incremental Redundancy

ISD Inter Site Distance

l.h.s. left hand side

LTE Long Term Evolution

LOS Line of Sight

M2M Machine-to-Machine

MDP Markov Decision Process

MDT Minimization of Drive Tests

MIMO Multiple-Input Multiple-Output

NACK Non Acknowledgment

NLOS Non-Line of Sight

NR New Radio

OPEX Operational Expenditure

OTDOA Observed Time Difference Of Arrival

ProSe Proximity Service

PS Public Safety

RAN Radio Access Network

RB Resource Block

RFID Radio Frequency Identification

r.h.s. right hand side

RSS Received Signal Strength

SINR Signal to Interference Noise Ratio

SMS Short Message Service

SNR Signal to Noise Ratio

TDD Time Division Duplexing

TDMA Time Division Multiple Access

ToA Time of Arrival

TTI Time Transmission Interval

UE User Equipment

UMTS Universal Mobile Telecommunications System

URLLC Ultra-Reliable and Low Latency communication

UWB Ultra-WideBand

V2I Vehicular-to-Infrastructure

V2N Vehicular-to-Network

V2P Vehicular-to-Pedestrian

V2V Vehicular-to-Vehicular

V2X Vehicular-to-Everything

VSC Virtual Small Cell

WLAN Wireless Local Area Network

WPAN Wireless Personal Area Network

7.5 Résumé en français

Titre: Intérêt de la Communication Directe entre Equipements Mobiles dans les Réseaux Radio sans fil

Dans le contexte des nouvelles générations de réseaux sans fil, la Cinquième-Génération (5G) est attendue dès 2020. Alors que l'objectif de la Quatrième-Génération (4G) était surtout de garantir les exigences requises par l'augmentation d'un débit maximum, le réseau 5G pourrait être vu comme une accumulation ou une mise en commun de plusieurs logiciels dans le but de proposer un grand nombre de services divers et variés. D'un point de vue technique, le réseau 5G devra être en mesure de répondre à trois grandes catégories de services : l'amélioration du haut débit mobile – *enhanced Mobile BroadBand (eMBB)* –, l'internet des objets massif – *massive Internet of Thing (IoT)* – et la communication ultra fiable et de latence très faible – *Ultra-Reliable and Low Latency communication (URLLC)*. Comme décrit sur la Fig. 1.1 [2]. L'eMBB est la continuité des services multimédia fournis par la 4G, pour lesquels les besoins concernent en particulier les hauts débits et l'extension de couverture radio. Dans le but de répondre à ces besoins, des technologies comme l'agrégation de porteuses, les nouvelles bandes centimétriques/millimétriques et la massive multi-entrée multi-sortie – *massive Multiple-Input Multiple-Output (MIMO)* – sont proposées parmi d'autres. Concernant l'internet des objets massif, l'omniprésente couverture radio avec grande densité d'équipements mobiles bon marché est proposée pour des services comme les villes intelligentes – *smart cities*. Pour l'URLLC, les contraintes en fiabilité et en faible latence sont sans précédent, et rendent possibles des services comme les voitures connectées, l'industrie 4.0, l'*e*-santé (santé connectée), les *smart cities* ou la réalité virtuelle parmi tant d'autres.

Tous les scénarios cités précédemment impliquent de nouvelles demandes et nous donnent l'opportunité de faire face à différents défis comme le grand nombre d'équipements mobiles connectés, les grandes couvertures radios, le streaming vidéo, la faible latence, la haute fiabilité, la précision de positionnement dans le cas intra-muros – *indoor* – et extérieurs – *outdoor*.

Une des technologies envisagées pour répondre à tous les défis cités précédemment et qui fait partie du réseau 5G est la communication directe entre équipements mobiles – *Device-to-Device (D2D) communication*. Cette technologie permet une communication directe entre deux terminaux mobiles, ce qui ajoute un poids capital au commerce de la 5G. En effet, la communication D2D pourrait être pour un opérateur téléphonique, un moyen de décharger le trafic du réseau. La com-

munication directe permet des services de proximité – *Proximity Service (ProSe)* – utiles pour les véhicules connectés (URLLC), les réseaux denses (massive IoT), mais aussi pour les extensions de couverture (eMBB). Ceci améliore l'utilisation spectrale, le débit moyen, la latence, et peut-être utilisé pour des applications qui utilisent la localisation.

En particulier, dans cette thèse, nous explorons plusieurs scénarios faisant intervenir des réseaux sans fil, tel que ceux représentés dans la Fig. 1.2, dans lesquels le D2D apporte une valeur ajoutée fondamentale. Les descriptions de ces scénarios et les contributions apportées sont résumées dans les paragraphes ci-dessous.

Scénario de multidiffusion dans les réseaux radio sans fil :

Un premier cas d'usage important dans les réseaux sans fil est le canal multidiffusion, dans lequel un seul émetteur envoie un message commun à plusieurs récepteurs à travers un canal qui s'estompe – *fading channel*. Le taux de la transmission multidiffusion peut être analysé de deux façons :

- Le taux moyen de multidiffusion, qui correspond au nombre moyen de bits décodés avec succès par canal utilisé pour un utilisateur choisi uniformément au hasard.
- Le taux de coupure - *outage* - de multidiffusion, qui correspond au taux maximum tel que tous les utilisateurs puissent décoder le message avec une probabilité d'erreur de l'ordre de ϵ

Tout d'abord, le taux moyen de multidiffusion est intéressant dans le contexte de la mise en mémoire cache sans fil – *wireless caching*. Il a été prouvé que le trafic durant les heures de pointe peut être considérablement réduit en mettant en mémoire cache les contenus répandus durant les heures creuses, et en distribuant ces contenus par multidiffusion [3, 4]. Le problème de ce taux moyen de multidiffusion est qu'il reste constant dans les scénarios denses, c'est-à-dire quand le nombre d'utilisateurs augmente.

Ensuite, l'étude du taux de coupure de multidiffusion est essentielle pour les scénarios comme l'amélioration des services multimédias par multidiffusion – *enhanced Multimedia Broadcast Multicast Service (eMBMS)* – [5]. Ces services réduisent la charge du réseau en multidiffusant une donnée commune aux équipements mobiles de sécurité publique [6] – *Public Safety (PS)* – pour lesquels il est impératif que chacun puisse décoder. Dans ce cas, le taux de coupure est limité par l'équipement dans la pire condition radio. On obtient donc un taux de coupure qui diminue

quand le nombre d'équipements augmente. Motivés par ces deux scénarios, nous allons étudier la multidiffusion sans fil aidée par la technologie D2D sans connaissance du canal à l'émetteur – *Channel State Information at the Transmitter (CSIT)* – pour surmonter l'effet de non-croissance du taux de multidiffusion. Plutôt que de considérer le CSIT, nous ne supposons uniquement qu'une connaissance statistique de cannal, qui est transmise par les récepteurs à chaque période de temps. Ceci évite une surcharge au moment d'estimer le cannal, ainsi que d'éventuelles erreurs liées à cette estimation dans le cas typique d'évanouissement rapide du canal. Nous considérons donc une topologie générale du réseau combiné avec la technologie D2D, où l'émetteur n'a qu'une connaissance statistique du canal. De plus, l'analyse du taux de multidiffusion s'effectuera dans un réseau dense, c'est-à-dire dans un régime où le nombre d'équipements mobiles est grand. Nous répondons donc à la question fondamentale : le D2D sans CSIT peut-il augmenter le taux de multidiffusion ?

Nos contributions dans ce scénario de multidiffusion sont les suivantes :

1 Nous proposons un procédé double-phases élémentaire qui atteint un taux de multidiffusion évolutif en fonction du nombre d'équipements avec une connaissance seulement statistique du canal à l'émetteur. En considérant une répartition optimale des ressources, ce procédé de double-phase amélioré atteint un taux moyen de multidiffusion de l'ordre de $\mathcal{O}(\ln \ln n)$ avec grande probabilité, où n représente le nombre d'équipements mobiles dans le réseau.

2 Nous déduisons une expression asymptotique traitable pour les deux types de taux de multidiffusion (moyen et de coupure) dans un régime asymptotique, où n est grand. Il en va de soi que ces expressions sont nouvelles et basées sur des inégalités de concentration non-triviales.

3 Nous fournissions par la suite une expression approximative pour les deux types de taux de multidiffusion. Nos exemples numériques montrent que ces expressions sont très précises même pour un faible nombre d'équipements mobiles.

Scénario d'URLLC dans les réseaux radio sans fil :

Un deuxième cas d'usage important est celui de la transmission URLLC. Ces cas d'usage ont besoin de fournir une transmission ultra fiable avec une latence très faible, tandis que ces deux contraintes sont généralement très conflictuelles.

Combiner l'URLLC et la technologie D2D permet d'introduire des services comme les voitures connectées (qui utilisent une extension du D2D appelée véhicule à véhicule – *Vehicular-to-Vehicular (V2V)* – ou encore véhicule à tout – *Vehicular-to-Everything (V2X)* –), mais aussi des services comme l'*e*-santé ou l'industrie 4.0 au moment de connecter les machines ensemble, c'est-à-dire la communication machine à machine – *Machine-to-Machine (M2M) communication*.

Sous la contrainte de latence de l'URLLC, une date butoir des paquets apparaît et les créneaux – *slots* – pour la transmission doivent être courts. C'est pourquoi l'organisation de projet *3rd Generation Partnership Project (3GPP)* a commencé à standardiser les créneaux de transmission de très courte durée (à titre de comparaison, 1 ms pour l'évolution long terme – *Long Term Evolution (LTE)* – utilisée dans la 4G, 500 μs et 125 μs pour la "nouvelle radio" – *New Radio (NR)* – pour le futur réseau 5G) [7, 8]. Ceci est dû aux espacements de sous-porteuses – *sub-carrier spacing* – plus larges que pour le LTE, possible grace à des bandes passantes – *bandwidth* – plus larges (20 MHz à une fréquence centrale de 700 MHz pour le LTE, 100 MHz à une fréquence centrale de 3.5 GHz et 400 MHz à 28 GHz pour la NR, respectivement par rapport aux créneaux de transmission [7]).

Maintenant, dans le but de faire face à la forte contrainte en fiabilité de l'URLLC, une procédure de retransmission de données comme celle de la demande automatique de répétition hybride – *Hybrid Automatic Repeat Request (HARQ)* – est obligatoire (en effet, dans le but d'atteindre une très haute fiabilité en présence de bruit, il est impossible d'éviter les retransmissions). Lors d'une communication D2D utilisant la procédure de retransmission HARQ, l'émetteur attend l'accusé de non-réception – *Non Acknowledgment (NACK)* – de la part du récepteur, avant de retransmettre son paquet erroné. Ce retard introduit donc de la latence, ce qui n'est ni adéquat ni négligeable, surtout avec des durées de créneaux aussi courtes que celles mentionnées précédemment qui sont nécessaires pour les cas d'usage exprimés. En effet, attendre le retour d'information – *feedback* – pourrait avoir comme conséquence que le paquet ne soit pas retransmis assez de fois au récepteur avant la date butoir, annulant les exigences de l'URLLC. En ce sens, une approche qui combine "attendre le NACK avant de retransmettre" et "retransmission aveugle" nécessite donc d'être implémentée avec la déduction d'une politique de retransmission optimale. Autrement dit, nous allons étudier, sous les contraintes exigeantes en matière de fiabilité et de latence, une approche qui consiste à retransmettre de façon optimale le même paquet sans nécessairement attendre le NACK.

Nos contributions dans ce scénario d'URLLC sont les suivantes :

1 Nous analysons un procédé d'HARQ optimal pour les services URLLC quand le retour d'information arrive avec un retard plus grand que la durée du créneau de transmission (comme c'est souvent le cas). Nous en déduisons la politique de retransmission optimale et démontrons qu'elle peut être calculée et implémentée en un temps raisonnable.

2 Nous proposons une nouvelle politique de retransmission semi-gloutonne – *semi-greedy* – qui a l'avantage d'être simple et qui ne dépend que de la connaissance des exigences des services (en terme de fiabilité et latence requises) et du retard provoqué par le temps nécessaire au retour d'information dans un canal. Cette politique est cependant optimale dans beaucoup de cas pratiques, que nous avons identifiés.

3 Nous étendons ensuite nos résultats au cas où plusieurs émetteurs, qui partagent le même canal sans fil, veulent transmettre leur paquet à un seul et même récepteur, toujours dans les conditions URLLC. Nous en déduisons une politique de retransmission qui fait face aux contraintes de l'URLLC avec le moins de retransmission possible.

Alors que les solutions proposées s'appliquent à beaucoup de cas d'usage de l'URLLC (qui incluent les transmissions émetteur vers stations de bases – *Base Stations (BSs)* –), elles sont particulièrement pertinentes pour la communication D2D sous contrainte URLLC. Cette dernière rend l'allocation de ressource difficile et implique des procédures telles qu' "écouter avant de parler". Dans ce cas précis, les NACKs pourraient être implicites, c'est-à-dire que l'absence d'accusé de bonne réception – *Acknowledgment (ACK)* – pourrait être interprétée comme un NACK vu que le récepteur pourrait ne pas être au courant de l'existence du paquet.

Scénario de localisation dans les réseaux radio sans fil :

Pour finir, dans le contexte du *massive IoT*, de nouveaux services apparaissent comme les *smart cities*. En général, dans de tels scénarios, les équipements interagissent les uns avec les autres, en situations *indoor/outdoor* qui nécessitent une très haute précision. La précision de l'emplacement est également très importante pour les services comme les voitures automatiques (éviter les accidents), l'*e*-santé (chirurgie ou encore assistance à domicile), les maisons intelligentes (domotique),

les réseaux sociaux (trouver des amis proches) et l'industrie 4.0 (entrepôts connectés). Remarquons que les équipements utilisés dans les *smart cities* ou industrie 4.0, également appelés "équipements intelligents", doivent être bon marché, à faible consommation énergétique et la transmission entre ces équipements doit être robuste par rapport à l'interférence cellulaire et à l'effet d'évanouissement du canal. Sachant qu'elle satisfait toutes les précédentes contraintes et qu'elle a été standardisée par le groupe de travail *Institute of Electrical and Electronics Engineers (IEEE)* 802.15, qui définit les standards pour les réseaux personnels sans fil – *Wireless Personal Area Network (WPAN)* –, nous proposons de considérer l'utilisation de signaux à ultra large bande – *Ultra-WideBand (UWB)* – pour le D2D. L'UWB propage de faibles pulsations énergétiques à travers une très large bande de fréquence, ce qui permet une très haute précision.

Vu que les équipements appartenant aux *smart cities* ou aux industries 4.0 ont tendance à être confinés dans des groupes locaux, il est d'autant plus intéressant pour le réseau local de connaître leur localisation et de les regrouper avec la plus grande précision possible. La localisation coopérative entre les équipements est donc une fonctionnalité clef pour le réseau 5G.

Nos contributions dans ce scénario de localisation sont les suivantes :

1 Nous proposons une méthode dynamique de coopération entre les équipements pour les cartographier, où la précision est étudiée pour les cas de propagation de canal avec et sans visibilité directe – *Line of Sight (LOS) and Non-Line of Sight (NLOS)* –, fournissant une précision de l'ordre du centimètre.

2 Nous proposons également une comparaison de la précision de notre méthode de localisation avec celle du *Global Positioning System (GPS)*.

3 Nous appliquons notre méthode de localisation entre équipements afin de détecter les zones denses – *hotspots*. De plus, dans le but d'éviter le déploiement d'infrastructure pour les petites cellules – *small cells* – également appelées point d'accès à un réseau, un grand nombre d'antennes côté BS peuvent être utilisées, afin de créer des faisceaux – *beam* – très directifs qui couvrent une zone donnée. Ce concept est appelé les petites cellules virtuelles – *Virtual Small Cells (VSCs)*. Ce scénario de couverture sera très utile pour les services 5G tels que les *smart cities*, l'industrie 4.0 ou encore les scénarios de couverture générale à bas coût.

Bibliography

[1] N. Mandela, *Long Walk to Freedom*. Macdonald Purnell, 1994.

[2] B. Sayrac, "Introduction to 5G: General context, services and requirements, new business models," Orange Labs, Tech. Rep., 2018.

[3] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.

[4] G. Paschos, E. Bastug, I. Land, G. Caire, and M. Debbah, "Wireless caching: Technical misconceptions and business barriers," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16–22, 2016.

[5] "3GPP TR 23.768 Study on Architecture Enhancements to Support Group Communication System Enablers for LTE (GCSE LTE)," 3GPP organization, Tech. Rep., Nov 2013.

[6] Nokia, "LTE networks for public safety services, white paper, 2016."

[7] R. Visoz, "RAN network: RAN functionalities and design," Orange Labs, Tech. Rep., 2018.

[8] "Designing 5G NR. The 3GPP Release-15 global standard for a unified, more capable 5G air interface," Qualcomm, Tech. Rep., April 2018.

[9] A. Galindo-Serrano, S. M. Lopez, A. De Ronzi, and A. Gati, "Virtual Small Cells Using Large Antenna Arrays as an Alternative to Classical HetNets," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, 2015, pp. 1–6.

[10] M. J. Yang, S. Y. Lim, H. J. Park, and N. H. Park, "Solving the data overload: Device-to-device bearer control architecture for cellular data offloading," *IEEE Vehicular Technology Magazine*, vol. 8, no. 1, pp. 31–39, 2013.

[11] "Technical specification group, feasability study for proximity services (ProSe), release 12, TR 22.803 v1.0.0," 3GPP organization, Tech. Rep., August 2012.

[12] "3rd Generation Partnership Project: Technical specification group core network and terminals; proximity-services user equipment to ProSe function protocol aspects; (release 13)," March 2014.

[13] "Technical Specification Group Radio Access Network; Study on LTE-based V2X Services, release 14, TR 22.885 v14.0.0," 3GPP organization, Tech. Rep., June 2016.

[14] "Technical Specification Group Services and System Aspects; Study on enhancement of 3GPP Support for 5G V2X Services, release 15, TR 22.886 v15.2.0," 3GPP organization, Tech. Rep., June 2018.

[15] K. Huang, V. K. Lau, and Y. Chen, "Spectrum sharing between cellular and mobile ad hoc networks: transmission-capacity trade-off," *IEEE Journal on selected Areas in Communications*, vol. 27, no. 7, 2009.

[16] B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum sharing scheme between cellular users and ad-hoc device-to-device users," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1038–1049, 2013.

[17] E. Karapistoli, F.-N. Pavlidou, I. Gragopoulos, and I. Tsetsinas, "An overview of the IEEE 802.15. 4a standard," *IEEE Communications Magazine*, vol. 48, no. 1, 2010.

[18] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, 2012.

[19] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Communications*, vol. 19, no. 3, 2012.

[20] P. Janis, Y. Chia-Hao, K. Doppler, C. Ribeiro, C. Wijting, H. Klaus, O. Tirkkonen, and V. Koivunen, "Device-to-device communication underlaying cellular communications systems," *International Journal of Communications, Network and System Sciences*, vol. 2, no. 03, p. 169, 2009.

[21] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 413–418, 2011.

[22] H. Min, J. Lee, S. Park, and D. Hong, "Capacity enhancement using an interference limited area for device-to-device uplink underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 3995–4000, 2011.

[23] K. Doppler, M. P. Rinne, P. Janis, C. Ribeiro, and K. Hugl, "Device-to-device communications; functional prospects for LTE-advanced networks," in *Communications Workshops, 2009. ICC Workshops 2009. IEEE International Conference on*. IEEE, 2009, pp. 1–6.

[24] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Transactions on Wireless communications*, vol. 10, no. 8, pp. 2752–2763, 2011.

[25] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[26] S. Mumtaz and J. Rodriguez, *Smart device to smart device communication*. Springer, 2014.

[27] S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina, and Y. Koucheryavy, "Cellular traffic offloading onto network-assisted device-to-device connections," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 20–31, 2014.

[28] B. Xing, K. Seada, and N. Venkatasubramanian, "An experimental study on wi-fi ad-hoc mode for mobile device-to-device video delivery," in *INFOCOM Workshops 2009, IEEE*. IEEE, 2009, pp. 1–6.

[29] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "3GPP LTE traffic offloading onto WiFi direct," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2013 IEEE*. IEEE, 2013, pp. 135–140.

[30] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.

[31] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Device-to-device collaboration through distributed storage," in *Global Communications Conference (GLOBECOM), 2012 IEEE*. IEEE, 2012, pp. 2397–2402.

[32] A. Asadi and V. Mancuso, "On the compound impact of opportunistic scheduling and D2D communications in cellular networks," in *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems*. ACM, 2013, pp. 279–288.

[33] Q. Wang and B. Rengarajan, "Recouping opportunistic gain in dense base station layouts through energy-aware user cooperation," in *2013 IEEE 14th International Symposium on*. IEEE, 2013, pp. 1–9.

[34] "3GPP TS 36.843 v12.0.1 study on LTE device to device proximity services; radio aspects (release 12)," 3GPP organization, Tech. Rep., March 2014.

[35] "On D2D communication related to out-of-coverage UE with TX timing not from eNB," 3GPP organization, Tech. Rep., August 2014.

[36] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, 2014.

[37] C.-H. Yu, O. Tirkkonen, K. Doppler, and C. Ribeiro, "Power optimization of device-to-device communication underlaying cellular communication," in *Communications, 2009. ICC'09. IEEE International Conference on*. IEEE, 2009, pp. 1–5.

[38] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, 2009.

[39] J. Gu, S. J. Bae, B.-G. Choi, and M. Y. Chung, "Dynamic power control mechanism for interference coordination of device-to-device communication in cellular networks," in *Ubiquitous and Future Networks (ICUFN), 2011 Third International Conference on*. IEEE, 2011, pp. 71–75.

[40] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for D2D communications," in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*. IEEE, 2011, pp. 358–362.

[41] S. Xu, H. Wang, T. Chen, T. Peng, and K. S. Kwak, "Device-to-device communication underlaying cellular networks: Connection establishment and interference avoidance." *KSII Transactions on Internet & Information Systems*, vol. 6, no. 1, 2012.

[42] C. Drula, C. Amza, F. Rousseau, and A. Duda, "Adaptive energy conserving algorithms for neighbor discovery in opportunistic bluetooth networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1, 2007.

[43] K. Stetson, "Wi-fi alliance announces groundbreaking specification to support direct wi-fi connections between devices," *Wi-Fi Alliance press release (PRNewswire)*, 2009.

[44] S. Martinez Lopez, "An overview of D2D in 3GPP LTE standards," *Indo-french workshop on D2D communication in 5G and IoT networks*, June 2016.

[45] F. Baccelli, N. Khude, R. Laroia, J. Li, T. Richardson, S. Shakkottai, S. Tavildar, and X. Wu, "On the design of device-to-device autonomous discovery," in *Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on*. IEEE, 2012, pp. 1–9.

[46] A. Vigato, L. Vangelista, C. Measson, and X. Wu, "Joint discovery in synchronous wireless networks," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2296–2305, 2011.

[47] K. Doppler, C. B. Ribeiro, and J. Kneckt, "Advances in D2D communications: Energy efficient service and device discovery radio," in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on*. IEEE, 2011, pp. 1–6.

[48] Alcatel-Lucent, "Discussion of D2D discovery methods," in *3GPP doc. R1-132068, RAN1, meeting 73, Fukuoka, Japan*. 3GPP, May 2013.

[49] "3GPP Technical Specification Group Core Network and Terminals; Proximity-services (ProSe) function to ProSe application server aspects (PC2); Stage 3, TS 29.343 V13.4.0 (Release 13)," 3GPP, Tech. Rep., Sept. 2016.

[50] S. Jung and J. Kim, "A new way of extending network coverage: Relay-assisted D2D communications in 3GPP," *ICT Express*, vol. 2, no. 3, pp. 117–121, 2016.

[51] M. J. Lopez, "Multiplexing, scheduling, and multicasting strategies for antenna arrays in wireless networks," MIT, Tech. Rep., 2004.

[52] P. K. Gopala and H. El Gamal, "On the throughput-delay tradeoff in cellular multicast," in *IEEE WNCMC*, 2005.

[53] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, 2006.

[54] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicast," in *Information Theory, 2006 IEEE International Symposium on*. IEEE, 2006, pp. 1841–1845.

[55] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[56] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

[57] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[58] S.-E. Elayoubi, A. M. Masucci, J. Roberts, and B. Sayrac, "Optimal D2D Content Delivery for Cellular Network Offloading," *Mobile Networks and Applications*, vol. 22, no. 6, pp. 1033–1044, 2017.

[59] K.-H. Ngo, S. Yang, and M. Kobayashi, "Scalable Content Delivery with Coded Caching in Multi-Antenna Fading Channels," *IEEE Trans. on Wireless Communications*, vol. 17, no. 1, pp. 548–562, 2018.

[60] R. Combes, A. Ghorbel, M. Kobayashi, and S. Yang, "Utility Optimal Scheduling for Coded Caching in General Topologies," *arXiv preprint arXiv:1801.02594*, 2018.

[61] "3GPP Technical Specification Group Services and System Aspects; Group Communication System Enablers for LTE (GCSE LTE); Stage 2, TS 23.468 V12.7.0 (Release 12)," 3GPP, Tech. Rep., Dec. 2015.

[62] X. Lin, R. Ratasuk, A. Ghosh, and J. G. Andrews, "Modeling, analysis, and optimization of multicast device-to-device transmissions," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4346–4359, 2014.

[63] D. Wang, X. Wang, and Y. Zhao, "An interference coordination scheme for device-to-device multicast in cellular networks," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*. IEEE, 2012, pp. 1–5.

[64] C.-W. Yeh, G.-Y. Lin, M.-J. Shih, and H.-Y. Wei, "Centralized interference-aware resource allocation for device-to-device broadcast communications," in *Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE*. IEEE, 2014, pp. 304–307.

[65] B. Peng, T. Peng, Z. Liu, Y. Yang, and C. Hu, "Cluster-based multicast transmission for device-to-device (D2D) communication," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th.* IEEE, 2013, pp. 1–5.

[66] J. Du, W. Zhu, J. Xu, Z. Li, and H. Wang, "A compressed HARQ feedback for device-to-device multicast communications," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE.* IEEE, 2012, pp. 1–5.

[67] J. Seppälä, T. Koskela, T. Chen, and S. Hakola, "Network controlled device-to-device (D2D) and cluster multicast concept for LTE and LTE-A networks," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE.* IEEE, 2011, pp. 986–991.

[68] B. Zhou, H. Hu, S.-Q. Huang, and H.-H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization," *IEEE transactions on vehicular technology*, vol. 62, no. 5, pp. 2315–2326, 2013.

[69] A. Khisti, U. Erez, and G. W. Wornell, "Fundamental limits and scaling behavior of cooperative multicasting in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2762–2770, 2006.

[70] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 146–151.

[71] 3GPP, "Study on scenarios and requirements for next generation access technologies," 3GPP TR 38.913 v14.2.0, Tech. Rep., June 2017.

[72] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition.* Wiley Publishing, 2011.

[73] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links," *IEEE Transactions on wireless communications*, vol. 3, no. 5, pp. 1746–1755, 2004.

[74] A. Galindo-Serrano, R. Visoz, M. Deghel, and S. E. El Ayoubi, "Joint Optimization of Link Adaptation and HARQ Retransmissions for URLLC Services," in *International Conference on Telecommunications*, 2018.

[75] 3GPP, "Physical layer procedures for data," 3GPP TR 38.214 v15.1.0, Tech. Rep., March 2018.

[76] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-based access for ultra-reliable low latency uplink transmissions," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 182–185, April 2018.

[77] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, February 2017.

[78] 3GPP, "Study on latency reduction techniques for LTE," 3GPP TR 36.881 v14.0.0, Tech. Rep., June 2016.

[79] "3GPP TR 36.885 v14.0.0 study on LTE-based V2X services; (release 14)," 3GPP organization, Tech. Rep., June 2016.

[80] "3GPP TR 36.886 v15.2.0 study on enhancement of 3GPP support for 5G V2X services; (release 15)," 3GPP organization, Tech. Rep., June 2018.

[81] "3GPP TR 36.886 v16.0.0 study on enhancement of 3GPP support for 5G V2X services; (release 16)," 3GPP organization, Tech. Rep., June 2018.

[82] "Exploring the world of wireless RAN meeting 80; 5G standardization update; vol. 14, no 6," Signals Ahead, Tech. Rep., June 2018.

[83] A. Galindo-Serrano, S. El Ayoubi, B. Sayrac, B. Graves, and A. De Lannoy, "Ultra reliable and low latency communication use cases and their requirements," in *Ambient connectivity/RAN design and critical communications*. 5G Radio System Design project, 11 2017.

[84] N. Brahmi, O. N. Yilmaz, K. W. Helmersson, S. A. Ashraf, and J. Torsner, "Deployment strategies for ultra-reliable and low-latency communication in factory automation," in *Globecom Workshops (GC Wkshps), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[85] "Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14, tr 38.913," 3GPP organization, Tech. Rep., Dec. 2016.

[86] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of Cellular Mobile Radio Localization Methods: from 1G to 5G," *IEEE Communications Surveys & Tutorials*, 2017.

[87] F. H. Blecher, "Advanced mobile phone service," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 2, pp. 238–244, 1980.

[88] G. D. Ott, "Vehicle location in cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 26, no. 1, pp. 43–46, 1977.

[89] M. Hata and T. Nagatsu, "Mobile location using signal strength measurements in a cellular system," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 2, pp. 245–252, 1980.

[90] J. H. Reed, K. J. Krizman, B. D. Woerner, and T. S. Rappaport, "An overview of the challenges and progress in meeting the E-911 requirement for location service," *IEEE Communications Magazine*, vol. 36, no. 4, pp. 30–37, 1998.

[91] H. Holma and A. Toskala, *WCDMA for UMTS: Radio access for third generation mobile communications.* john Wiley & sons, 2005.

[92] T. Farley, "Mobile telephone history," *Telektronikk*, vol. 101, no. 3/4, p. 22, 2005.

[93] "Report and order and further notice of proposed rulemaking on revision of the FCC rules to ensure compatibility with enhanced 911 emergency calling systems," Federal Commun. Commission, Washington, DC, USA, Rep. 96-264, Tech. Rep., 1996.

[94] "UE positioning enhancements TR 25.847 V1.0.0 (Release 4)," 3GPP Sophia Antipolis, France, Tech. Rep., Dec. 2000.

[95] "Stage 2 functional specification of UE positioning in UTRAN, TS 25.305 V7.0.0 (Release 7)," 3GPP Sophia Antipolis, France, Tech. Rep., Jun. 2005.

[96] "Stage 2 functional specification of UE positioning in UTRAN, TS 25.305 V10.0.0 (Release 10)," 3GPP Sophia Antipolis, France, Tech. Rep., Sep. 2010.

[97] "Way forward on OTDOA positioning," 3GPP Sophia Antipolis, France, Tech. Rep., Mar. 2009.

[98] S. Fischer, "Observed time difference of arrival (OTDOA) positioning in 3GPP LTE," *Qualcomm White Pap*, vol. 1, no. 1, pp. 1–62, 2014.

[99] "Study on indoor positioning enhancements for UTRA and LTE, TR 37.857 V13.1.0 (Release 13)," 3GPP Sophia Antipolis, France, Tech. Rep., Jan. 2016.

[100] "3GPP system to WLAN interworking; system description, TS 23.234 V12.0.0 (Release 12)," 3GPP Sophia Antipolis, France, Tech. Rep., Sep. 2014.

[101] "3GPP TS 36.213 Physical layer procedures (Release 12)," 3GPP organization, Tech. Rep., 2011.

[102] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[103] J.-W. Qiu, C. C. Lo, C.-K. Lin, and Y.-C. Tseng, "A D2D Relative Positioning System on Smart Devices," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 2168–2172.

[104] C. Fretzagias and M. Papadopouli, "Cooperative location-sensing for wireless networks," in *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on.* IEEE, 2004, pp. 121–131.

[105] L. T. Nguyen, S. Kim, and B. Shim, "Localization in the internet of things network: A low-rank matrix completion approach," *arXiv preprint arXiv:1702.04054*, 2017.

[106] J. Zhang, P. V. Orlik, Z. Sahinoglu, A. F. Molisch, and P. Kinney, "UWB systems for wireless sensor networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 313–331, 2009.

[107] W. C. Chung and D. Ha, "An accurate ultra wideband (UWB) ranging for precision asset location," in *Ultra Wideband Systems and Technologies.* IEEE, 2003, pp. 389–393.

[108] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis," *IEEE Transactions on Communications*, vol. 58, no. 4, 2010.

[109] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities. A nonasymptotic theory of independence.* Oxford University Press, 2013.

[110] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons, 2014.

[111] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Transactions on Communications*, vol. 25, no. 1, pp. 117–128, 1977.

[112] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on selected areas in communications*, vol. 18, no. 3, pp. 535–547, 2000.

[113] ERICSSON, "Positioning with LTE, maximizing performance through integrated solutions," *White Paper*, Sept 2011.

[114] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[115] T. Varela Santana, A. Galindo-Serrano, B. Sayrac, and S. Martínez López, "Dynamic network configuration: Hotspot identification for Virtual Small Cells," in *IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2016, pp. 49–53.

[116] K. I. Kossonou, "Étude d'un système de localisation 3-D haute précision basé sur les techniques de transmission ultra large bande à basse consommation d'énergie pour les objets mobiles communicants." Ph.D. dissertation, 2014.

[117] M. S. Svalastog, "Indoor positioning-technologies, services and architectures," Master's thesis, 2007.

[118] R. Mautz, "Indoor positioning technologies," 2012.

[119] K. Witrisal, S. Hinteregger, J. Kulmer, E. Leitinger, and P. Meissner, "High-Accuracy Positioning for Indoor Applications: RFID, UWB, 5G, and beyond," in *IEEE International Conference on RFID*, 2016, pp. 1–7.

[120] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal processing magazine*, vol. 22, no. 4, pp. 54–69, 2005.

[121] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE signal processing magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[122] S. Al-Jazzar and J. Caffery, "ML and Bayesian TOA location estimators for NLOS environments," in *IEEE 56th Vehicular Technology Conference. Proceedings. VTC 2002-Fall*, vol. 2, pp. 1178–1181.

[123] S. Al-Jazzar, J. Caffery, and H.-R. You, "A scattering model based approach to NLOS mitigation in TOA location systems," in *IEEE 55th Vehicular Technology Conference. Proceedings. VTC 2002-Spring*, vol. 2, pp. 861–865.

[124] B. Denis, J. Keignart, and N. Daniele, "Impact of NLOS propagation upon ranging precision in UWB systems," in *Ultra Wideband Systems and Technologies, 2003 IEEE Conference on.* IEEE, 2003, pp. 379–383.

[125] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the Nodes: Cooperative localization in wireless sensor networks," *IEEE Signal processing magazine, 2005*, vol. 22, no. 4, pp. 54–69.

[126] A. F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15. 4a channel model-final report," *IEEE P802*, vol. 15, no. 04, p. 0662, 2004.

[127] A. F. Molisch, P. Orlik, Z. Sahinoglu, and J. Zhang, "UWB-based sensor networks and the IEEE 802.15. 4a standard-a tutorial," in *ChinaCom'06*. IEEE, 2006, pp. 1–6.

[128] J. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Cramér-Rao lower bounds for the synchronization of UWB signals," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 426–438, 2005.

[129] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter wave MIMO in 5G systems," *IEEE Transactions on Wireless Communications*, 2017.

[130] G. Fischer, O. Klymenko, D. Martynenko, and H. Luediger, "An impulse radio UWB transceiver with high-precision TOA measurement unit," in *IPIN*. IEEE, 2010, pp. 1–8.

[131] R. Merz, F. Chastellain, A. Blatter, C. Botteron, and P.-A. Farine, "An experimental platform for an indoor location and tracking system," in *ENC, Toulouse, France*, 2008.

[132] G. Cheng, "Accurate TOA-based UWB localization system in coal mine based on WSN," *Physics Procedia*, vol. 24, pp. 534–540, 2012.

[133] M. Segura, V. Mut, and C. Sisterna, "Ultra wideband indoor navigation system," *IET Radar, Sonar & Navigation*, vol. 6, no. 5, pp. 402–411, 2012.

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

[135] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37–43, 2011.

[136] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing—The Large-Scale Array Regime," *IEEE transactions on information theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[137] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, 2014.

[138] S. Fortes, A. Aguilar-García, R. Barco, F. B. Barba, J. A. Fernández-Luque, and A. Fernández-Durán, "Management architecture for location-aware self-organizing LTE/LTE-a small cell networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 294–302, 2015.

[139] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and Challenges with Very Large Arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[140] S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy Efficiency and Sum Rate Tradeoffs for Massive MIMO Systems with Underlaid Device-to-Device Communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 175, 2016.

[141] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 53–64. [Online]. Available: http://doi.acm.org/10.1145/2348543.2348553

[142] Y. Liu, X. Duan, G. Boudreau, A. Bin Sediq, and X. Wang, "Adaptive Beamforming Based Inband Fronthaul for Cost-Effective Virtual Small Cell in 5G Networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2017.

[143] V. G. Vassilakis, H. Mouratidis, E. Panaousis, I. D. Moscholios, and M. D. Logothetis, "Security requirements modelling for virtualized 5G small cell networks," in *24th IEEE International Conference on Telecommunications (ICT)*, 2017, pp. 1–5.

[144] A. Radwan, M. F. Domingues, and J. Rodriguez, "Mobile caching-enabled small-cells for delay-tolerant e-Health apps," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 103–108.

[145] Q. Zhang, J. Zeng, X. Su, L. Rong, and X. Xu, "Virtual Small Cell Selection Schemes Based on Sum Rate Analysis in Ultra-Dense Network," in *International Conference on Communicatins and Networking in China*. Springer, 2016, pp. 78–87.

[146] "3GPP TR 36.814 evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (release 9)," 3GPP organization, Tech. Rep., 2010.

[147] T. Varela Santana, S. Martínez López, and A. Galindo-Serrano, "The virtual small cells based on UE positioning: a network densification solution," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, p. 42, Jun 2018.

[148] A. Dammann, R. Raulefs, and S. Zhang, "On Prospects of Positioning in 5G," in *IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 1207–1213.

[149] "3GPP TS 37.320 v11.1.0 universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); radio measurement collection for minimization of drive tests (MDT); overall description; stage 2 (release 11)," 3GPP organization, Tech. Rep., Sept. 2012.

[150] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog, "Minimization of drive tests in 3GPP Release 11," *IEEE Communications Magazine*, vol. 50, no. 11, 2012.

[151] T. Kanungo, N. S. Netanyahu, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," in *IEEE transactions on pattern analysis and machine intelligence*, July 2002.

[152] D. T. Pham, S. S. Dlmov, and C. D. Nguyen, "Selection of K in K-means clustering," *Manufacturing Engineering Centre, Cardiff University, Cardiff, UK*, pp. 103–119, Sep. 2004.

[153] P. A. Zandbergen, "Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning," *Transactions in GIS*, vol. 13, no. s1, pp. 5–25, 2009.

[154] T. Varela Santana, R. Combes, and M. Kobayashi, "Device-to-Device aided Multicasting," *submitted to IEEE Transactions on Wireless Communications*, 2018.

[155] T. Varela Santana, R. Combes, and S. E. El Ayoubi, "On ARQ With Delayed Feedback under Latency and Reliability Constraints," *Submitted to IEEE INFOCOM*, 2019.

[156] T. Varela Santana, R. Combes, and M. Kobayashi, "Device-to-Device aided Multicasting," *IEEE International Symposium on Information Theory (ISIT)*, pp. 771–775, 2018.

[157] T. Varela Santana, S. Arreffag, and S. Martínez López, "A High Resolution Method for Equipment Group Mapping Using UWB Signals," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2017.

[158] T. Varela Santana and S. Martínez López, "Procédés et dispositifs de cartographie flottante d'un groupe de mobiles," Patent 1 752 999, 04 06, 2017.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Intérêt de la Communication Direct entre Equipements Mobiles dans les Réseaux Radio sans fil

Mots clés: D2D, 5G, multidiffusion, URLLC, HARQ, Positionnement

Résumé: Dans cette thèse, nous étudions plusieurs scénarios de communication pour les futurs réseaux sans fil. Plus particulièrement, cette thèse porte son attention sur *comment la communication directe entre équipements mobiles - Device-to-Device (D2D) - peut améliorer les performances des technologies existantes dans les systèmes sans fil.* Le premier scénario étudié durant cette thèse est celui de la communication par multidiffusion d'un message commun entre un émetteur et plusieurs récepteurs. Il peut être illustré par le streaming vidéo, les messages d'alerte à destination de la police ou des pompiers ou des ambulanciers. Le second scénario étudié est celui d'une transmission à contraintes critiques en latence et en fiabilité. Ce dernier est illustré par son implication primordiale dans les futures technologies telles que les voitures connectées, avec pour but d'éviter des accidents, ou bien les machines connectées pour améliorer les services hospitaliers tels que la télé-chirurgie entre autres. Le dernier scénario étudié est celui de la localisation d'un groupe d'équipement dans un réseau densément peuplé tel qu'on peut trouver dans le contexte des objets connectés en masse. En général les objets communiquent entre eux à un niveau local et sont intéressés par des services communs et locaux. Plus concrètement, dans cette thèse, nous montrons les bienfaits de la communication D2D dans les trois scénarios précédents. Dans le cas du premier scénario de multidiffusion, contrairement à la tendance habituelle d'avoir un taux de transmission qui diminue en fonction du nombre d'équipements mobiles (en particulier, car l'équipement émetteur doit adapter sa transmission à l'équipement récepteur en plus mauvaise condition), en ajoutant la communication D2D, on observe que ce même taux de transmission augmente en fonction du nombre d'équipements mobiles présents. Dans le deuxième scénario où la communication est soumise à des contraintes de fiabilité et de latence exigeantes, nous déduisons une politique de retransmission optimale et proposons une autre politique semi-optimale qui est beaucoup moins gourmande en temps et qui a prouvé son optimalité dans plusieurs cas pratiques. Enfin dans le dernier scénario, nous proposons une méthode de localisation d'équipements mobile et l'étudions dans plusieurs environnements (avec et sans visibilité directe dans les cas intra-muros et extérieurs). L'identification de ces zones est ensuite utilisée pour créer de petites cellules virtuelles adaptatives aux situations changeantes et non prédictibles, dans le but de réduire les coûts liés aux infrastructures actuelles.

Title: On the use of Device-to-Device in Wireless Networks

Keywords: D2D, 5G, multicasting, URLLC, HARQ, Positioning

Abstract: This thesis studies Device-to-Device (D2D) communication in realistic and challenging scenarios for future wireless systems. In particular, the thesis focuses on *how may D2D communication help other technologies to enhance their performance.* The first wireless scenario is the one of multicasting, used for example in video streaming or common alert message transmission for police, firefighters or ambulances. The second wireless scenario is the critical one of Ultra-Reliable and Low Latency communication (URLLC) expected to be used to avoid cars crashes in the upcoming Vehicular-to-Everything (V2X) context, and also when connecting machines together in environments like connected hospitals, airports, factories (industry 4.0), and last but not least in e-health context in order to enhance medical tele-surgery. The last wireless scenario is the one of User Equipment (UE) group localization in the context of massive Internet of Thing (IoT), where devices are interacting with each other and are mostly confined in local groups, needing local services. In the multicast channel scenario, where a transmitter wishes to convey a common message to many receivers, it is known that the multicast rate decrease as the number of UEs increases. This vanishing behavior changes drastically when enabling the receivers to cooperate with each other via D2D. Indeed, the multicast rate increases with high probability when the number of receivers increases. This chapter also analyzes the outage rate of the proposed scheme in the same setting. Extensions regarding firstly resource utilization and secondly considering the use of Hybrid Automatic Repeat Request (HARQ) are also analyzed. Next chapter addresses one of the major challenges for future networks, named URLLC. Specifically, the chapter studies the problem of HARQ with delayed feedback, where the transmitter is informed after some delay on whether or not his transmission was successful. The goal is to minimize the expected number of retransmissions subject to a reliability constraint within a delay budget. This problem is studied at two levels: (i) a single transmitter faced with a stochastic i.i.d. noisy environment and (ii) a group of transmitters whom shares a collision channel. Then the chapter that follows provides a cooperative UE mapping method that is highly accurate. Four different channel models are studied in this chapter: Line of Sight (LOS) and Non-Line of Sight (NLOS) for indoor and outdoor environments. The results show significant improvement compared to already existing methods. Identifying the dense local areas in real time and informing the network allows the Base Station (BS) to increase the capacity through highly directive beams, and therefore, avoids the deployment cost of new infrastructure.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Rue Joliot-Curie 91190 Saint-Aubin, France