ENHANCED HANDOVER MECHANISMS FOR 5G MILLIMETER WAVE VEHICULAR NETWORKS

Davi da Silva Brilhante

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: José Ferreira de Rezende
Nicola Marchetti

Rio de Janeiro
Novembro de 2024

ENHANCED HANDOVER MECHANISMS FOR 5G MILLIMETER WAVE VEHICULAR NETWORKS

Davi da Silva Brilhante

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: José Ferreira de Rezende
Nicola Marchetti

Aprovada por: Prof. Daniel Ratton Figueiredo
Prof. Kleber Vieira Cardoso
Prof. Igor Monteiro Moraes

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2024

Brilhante, Davi da Silva

Enhanced Handover Mechanisms for 5G Millimeter Wave Vehicular Networks/Davi da Silva Brilhante. – Rio de Janeiro: UFRJ/COPPE, 2024.

XXI, 108 p.: il.; 29, 7cm.

Orientadores: José Ferreira de Rezende

Nicola Marchetti

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2024.

Referências Bibliográficas: p. 94 – 104.

1. Handover. 2. 5G New Radio. 3. Vehicular Networks. I. Rezende, José Ferreira de *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Dedicated to Maria José Alves da Silva (In memoriam)

Agradecimentos

Primeiramente minha gratidão é direcionada aquEle que existe antes de todas as coisas e subsistirá depois de todas as coisas.

Agradeço especialmente à minha esposa, Thayná. Não importa os obstáculos que surgiram diante de nós eu pude superá-los junto a ela e contando a sua parceria e amizade. Agora, ao fim dessa jornada, seu apoio foi essencial para o meu dia a dia, os quais você iluminou com sua presença gentil e delicada. Obrigado pela sua compreensão, pelas repetidas vezes que você me ouviu e por todos esses dias que você esteve ao meu lado, pois tudo teria sido muito mais difícil sem você aqui.

Junto a ela, estão também meus familiares, que acompaharam de perto a construção deste trabalho e me apoiaram a todo momento. Meus pais, sogros, irmãos e cunhados, são uma parte importante da minha formação e consequentemente deste trabalho. Eu sinceramente não imagine que pudesse chegar até aqui, mas se eu cheguei é devido às direções que vocês me influenciaram a tomar e ao seu apoio para perseverar em meio as adversidades.

A vida me presenteou também com leais amigos, que muito ajudaram a aliviar meus dias com suas estimadas companhias. Fábio David e Conrado Catarcione, representam meus amigos do laboratório LAND, que desde 2016 suportaram minha cômica personalidade rabugenta e me proporcionaram deliciosas conversas acompanhadas de café e biscoitos (muitas dessas interações resultaram em partes do texto que segue). Octávio Azevedo, Guilherme Scoffano, João Felipe Santos e Fadhil Firyaguna são amigos de outras épocas com quem de vez em quando eu posso dividir os questionamentos da vida e que me ajudam, mesmo que não saibam, a entender os picos e vales da estrada da vida. O parágrafo a seguir está escrito em inglês para garantir que todas as pessoas endereçadas nele possam entender.

Rezende and Nicola, you have been exemplary models of what a professor should be. I am fortunate to have had your support throughout this PhD journey. You provided the guidance I needed to keep tracking this path until the end. People cannot imagine the deepness of your patience, the extent of your understanding, and how much I learnt from you these years. Beyond being outstanding advisors, you are also incredible human beings, whom I profoundly admire. I hope that one day I can embody even a fraction of the character you both have shown me.

Por fim, muitos outros com certeza estiveram envolvidos no desenvolvimento desta tese. É difícil estimar a contribuição de tantos indivíduos que me influenciaram positivamente e pelos quais eu tenho tamanha estima. Citá-los nominalmente seria difícil e eu poderia cometer a falha de esquecer de mencionar alguém. Portanto, se temos carinho mútuo e em algum momento da minha nós sorrimos ou choramos juntos, deixo-lhe o meu mais sincero muito obrigado.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MECANISMOS OTIMIZADOS DE HANDOVER PARA REDES VEICULARES 5G EM ONDAS MILIMÉTRICAS

Davi da Silva Brilhante

Novembro/2024

Orientadores: José Ferreira de Rezende
Nicola Marchetti

Programa: Engenharia de Sistemas e Computação

Em redes de ondas milimétricas, devido ao raio de alcance limitado da portadora e aos feixes estreitos usados na transmissão, as técnicas de *handover* tradicionais podem não corresponder aos requisitos de aplicações de comunicação de alta confiabilidade e baixa latência (do inglês, *Ultra-Reliable Low-Latency Communication*, URLLC) e que também demandam enlaces de alta capacidade, como veículos autônomos, por exemplo. Nesta tese, é proposto um modelo de otimização para o estabelecimento de um limite superior de desempenho para a avaliação de mecanismos de *handover*. O modelo proposto é comparado aos mecanismos de *handover* da Quinta Geração de Comunicação Móvel (do inglês, *Fifth Generation of Mobile Communication*, 5G), chamados *baseline* e *conditional handovers*. Também propomos um mecanismo de handover baseado em uma heurística de programação dinâmica que explora densidade de bloqueios para tomar decisões de *handover*. Em termos de desempenho, a solução ótima proposta e a heurística são mais robustas que os mecanismos de *handover* do 5G na taxa de handovers realizados, taxa de *handovers* causados por ping-pong, e latência. Nossa avaliação indica que os mecanismos de *handover* do 5G pode ser aprimorados para que possam cumprir com os requisitos rigorosos das aplicações modernas para dispositivos móveis. Isto é evidenciado pela redução média no efeito ping-pong de 68% e 46%, e queda de 71% e 67% na taxa de *handovers* realizados se comparados à solução ótima e a heurística ao mecanismo de *conditional handover*, respectivamente.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ENHANCED HANDOVER MECHANISMS FOR 5G MILLIMETER WAVE VEHICULAR NETWORKS

Davi da Silva Brilhante

November/2024

Advisors: José Ferreira de Rezende
Nicola Marchetti

Department: Systems Engineering and Computer Science

In millimetre wave networks, due to the limited communication range and the narrow antenna beams, traditional handover techniques might not fully comply with the requirements of high-capacity Ultra-Reliable Low-Latency Communication (URLLC) applications, such as autonomous vehicles. In this thesis, we propose an optimal solution that sets performance boundaries for handover mechanisms.We compare the performance of the Fifth Generation of Mobile Communication (5G) handover mechanisms, called baseline and conditional handovers, with the proposed optimal solution in terms of the number of handovers, ping-pong effect ratio, and delay. We also propose a heuristic handover mechanism based on dynamic programming which exploits blockage predictions to make informed handover decisions. Our assessment indicates that the 5G handover mechanisms can be improved to meet the novel applications' tight constraints, evidenced by a ping-pong effect average reduction of 68% and 46%, and a decrease of 71% and 67% of the handover rate comparing the optimization and the heuristic handover mechanism to the conditional handover method, respectively.

Contents

List of Figures

List of Tables

List of Symbols

A_k^n	UE n k-th message arrival time, p. 37
$G(.)$	Antenna gain function, p. 58
G_{max}	Main lobe directional antenna gain, p. 58
G_{min}	Secondary lobe directional antenna gain, p. 58
$G_{rx,dB}$	Receiver antenna gain in dB, p. 37
$G_{tx,dB}$	Transmitter antenna gain in dB, p. 37
H	A3 Hysteresis margin, p. 22
K_n	Number of UE n data messages, p. 41
L	UE trajectory length, p. 62
M	Set of Base Stations, p. 57
N	Set of User Equipments, p. 57
N_{L1}	Number averaged Layer 1 measurement samples, p. 20
O	A3 Handover Offset, p. 22
$P_{L,dB}(d_{n,t}^m)$	Path Loss in dB between UE n and BS m at timeslot t, p. 37
$P_{tx,dB}^m$	BS transmission power, p. 37
Q_{in}	In-sync quality indicator, p. 23
Q_{out}	Out-of-sync quality indicator, p. 23
R_S	Serving BS RSRP, p. 22
R_T	Target BS RSRP, p. 22
R_{Th}	Heuristic handover BS blockage prediction request range, p. 44

SC_η	subcarrier spacing for the chosen numerology η, p. 39
T	Simulation observation period, p. 35
α	Layer 3 forgetting factor, p. 20
$\bar{K}_n$	Number of discarded packets, p. 66
δ	Transmission Time Interval duration, p. 57
η	5G NR chosen PHY Layer numerology, p. 39
η	Numerology Index, p. 17
γ	Packet Length, p. 63
λ	Blockage density, p. 62
λ_p	Packet arrival mean, p. 63
$\lfloor . \rfloor$	Floor Operation, p. 105
$\mathcal{B}_{\text{pred}}(m,n,t)$	Blockage prediction for BS m and the UE n at timeslot t, p. 47
$\mathcal{C}_{n,t}^{m}$	Shannon's theoretical capacity for the UE n associated to the BS m at timeslot t, p. 39
$\mathcal{C}_n$	UE n Capacity requirement, p. 37
$\mathcal{D}_n$	UE n delay tolerance, p. 41
$\mathcal{I}_{n,t}$	UE n delay penalization function at timeslot t, p. 41
$\mathcal{L}$	Blockage prediction length, p. 44, 47
$\mathcal{M}'$	Subset of candidate BSs selected by the heuristic handover algorithm, p. 47
$\mathcal{N}$	Gaussian random variable, p. 37
$\mathcal{O}(.)$	Worst-case computational complexity, p. 108
$\mathcal{P}_{n,t}^{m}$	Received power in dB, p. 37
$\mathcal{Q}$	Set of handover candidate BSs, p. 39
$\mathcal{R}_m$	BS m physical resource blocks, p. 37
$\mathcal{S}_{n,t}^{m}$	SNR between UE n and BS m at a given timeslot t, p. 37

$\mathcal{W}_k$	k-th message delay window, p. 41
μ	Gaussian random variable mean, p. 37
ρ	BS radius, p. 62
σ	Random variable standard deviation, p. 37
$\sigma^2_{\text{noise},dB}$	Additive white Gaussian noise power, p. 37
τ	Time to trigger, p. 37
θ_b	Antenna main lobe beam width, p. 58
φ	Physical Resource Block subcarrier bandwidth, p. 57
$d^m_{n,t}$	Distance between BS m and UE n at the timeslot t, p. 37
ell	Dynamic programming algorithm timestep, p. 47
k	Layer 3 filter coefficient exponent, p. 20
$kappa$	Out-of-synchronization penalization factor, p. 42
p_{ho}	Heuristic handover penalization, p. 47
t	timeslot, p. 40
t_{L1}	Layer 1 measurement sample index, p. 20
t_{L3}	Layer 3 measurement period, p. 20
t_{L3}	Layer 3 measurement sample index, p. 20
$u^p_{n,t}$	Handover opportunity binary indicator variable, p. 38
v	UE velocity in km/h, p. 62
$v^m_{n,t}$	SINR below out-of-synchronization threshold Q_{out} indicator variable, p. 42
$w^m_{n,t}$	Heuristic handover decision variable, p. 47
$x^m_{n,t}$	User equipment association decision variable, p. 35
$y^m_{n,t}$	Message delay decision variable, p. 40
GHz	Giga Hertz, p. 1
KHz	Kilo Hertz, p. 17

Mbps	Mega bits-per-second, p. 6
N310	Maximum number of consecutive out-of-sync indications, p. 23
N311	Minimum number of consecutive in-sync indications, p. 23
T304	Handover completion timer, p. 24
T310	Out-of-sync timer, p. 23
T301	RRC Connection Re-establishment response timer, p. 25
T311	Time to Initiate the RRC Connection Reestablishment Request, p. 24
a	Error function minimal error probability, p. 61
b	Error function exponent multiplier, p. 61
dB	Decibel, p. 14
handovers/s	Handovers per second, p. 78
km/h	Kilometer per hour, p. 62
ms	milli seconds, p. 1
obstacles/m^2	Obstacle per square-meter, p. 79

List of Abbreviations

3GPP	Third Generation Partnership Project, p. 1
4G	Fourth Generation of Mobile Communication, p. 15
5G	Fifth Generation of Mobile Communication, p. 1, 2
ARQ	Automatic Repeat Request, p. 24
AR	Augmented Reality, p. 16
Anatel	Brazilian national agency of telecommunication (*Agência Nacional de Telecomunicações*, in Portuguese), p. 1
BDMA	Beam Division Multiple Access, p. 16
BS	Base Station, p. 2
CA	Carrier Aggregation, p. 17
CF M-MIMO	Cell-Free Massive MIMO, p. 44
CN	Core Network, p. 7
CODEC	Codification Decodification, p. 54
CPU	Central Processing Unit, p. 44
CRAN	Cloud Radio Access Network, p. 32
CSI	Channel State Information, p. 2
D2D	Device-to-Device, p. 32
DDPG	Deep Deterministic Policy Gradient, p. 31
DMRS	DeModulation Reference Signal, p. 17
DNN	Deep Neural Network, p. 32
EMBB	Enhanced Mobile Broadband, p. 6

EWMA	Exponential Weighted Moving Average, p. 20
FPS	Frames Per Second, p. 63
FR	Frequency Range, p. 1
HIT	Handover Interruption Time, p. 6
HPPP	Homogeneous Poisson Point Process, p. 62
IIR	Infinite Impulse Response, p. 20
IP	Internet Protocol, p. 25
ITS	Intelligent Transportation System, p. 5, 16
IoT	Internet of Things, p. 16
KPI	Key Performance Indicators, p. 15
LIDAR	Light Detection and Ranging, p. 5
LP	Linear Programming, p. 29
LSTM	Long-Short Term Memory, p. 33
LTE	Long-Term Evolution, p. 1
LiDAR	Light Detection and Ranging, p. 32
LoS	Line of Sight, p. 2
MAB	Multi-Armed Bandit, p. 32
MDP	Markov Decision Process, p. 31
MID	Master Information Block, p. 25
MILP	Mixed-Integer Linear programming, p. 5
ML	Machine Learning, p. 10
MMTC	Massive Machine Type Communication, p. 16
MR	Message Report, p. 21
MSG	Message, p. 18
MToS	Minimal Time of Staying, p. 23

NBS	Neighbour BS, p. 26
NLoS	Non-Line of Sight, p. 2
NR	New Radio, p. 2
Near-RT-RIC	Near Real-Time RAN Intelligent Controller, p. 54
OFDM	Orthogonal Frequency Division Multiplexing, p. 16
OpenRAN	Open Radio Access Network, p. 54
PHY	Physical Layer, p. 13
PRB	Physical Resource Block, p. 57
PSS	Primary Synchronization Signal, p. 17
QoS	Quality of Service, p. 19
RACH	Random-Access Channel, p. 17
RADAR	Radio Detection and Ranging, p. 43
RAN	Radio Access Network, p. 19
RAT	Radio Access Technology, p. 18
RLC	Radio Link Control, p. 24
RLF	Radio Link Failure, p. 4, 23
RMSE	Root-Mean-Square Error, p. 29
RNA	RAN Notification, p. 19
RRC	Radio Resource Control, p. 18
RRM	Radio Resource Management, p. 22
RSRP	Reference Signal Received Power, p. 17
RSSI	Received Signal Strength Indicator, p. 33
SBS	Serving Base Station, p. 18
SCS	Subcarrier Spacing, p. 16
SETU	South East Technological University, p. 93

SIB1	System information Block message 1, p. 25
SINR	Signal-to-Interference-plus-Noise Ratio, p. 57
SNR	Signal-to-Noise Ratio, p. 31
SSS	Secondary Synchronization Signal, p. 17
SS	Synchronization Signal, p. 17
TBS	Target Base Station, p. 18
TTI	Transmission Time Interval, p. 57
TTT	Time to Trigger, p. 4, 11, 21
ToS	Time of Staying, p. 23
UE	User Equipment, p. 6
UMTS	Universal Mobile Telecommunication System, p. 18
URLLC	Ultra-Reliable and Low-Latency Communication, p. 1
V2I	Vehicle to Infrastructure, p. 6
V2V	Vehicle to Vehicle, p. 16
V2X	Vehicle to Everything, p. 2
VR	Virtual Reality, p. 16
WLAN	Wireless Local Area Network, p. 17
XGBoost	eXtreme Gradient Boosting, p. 29
mmWave	Millimeter Waves, p. 1

Chapter 1

Introduction

In September 2021, Anatel, the Brazilian national agency of telecommunication, had the biggest spectrum licensing auction in its history, the Fifth Generation of Mobile Communication (5G) auction summing up R$ 46.8 billion [6, 7]. Among the licensed frequencies, was the 26 GHz band, licensed by almost R$ 120 million to three service providers. All this interest in an invisible asset, such as the electromagnetic spectrum, reveals the general expectation of 5G. Indeed, the aspirations for 5G are high, and society, industry, and academia have put a lot of effort into making the evolution of Long-Term Evolution (LTE) to embrace a wide variety of applications with outstanding performance.

One class of applications of 5G, Ultra-Reliable and Low-Latency Communication (URLLC), poses the most challenging requirements. The URLLC services are characterized by latency below 1 ms and less than 10^{-5} outage probability [8]. Coping with such stringent requisites implies using advanced communication technologies and well-designed solutions. For example, congestion control and transport network delay mitigation are some areas to be improved for URLLC deployment [9]. Apart from that, some URLLC applications go beyond the already tough requirements and demand high throughput communication as well, like autonomous vehicles and remote driving. Nowadays, vehicles embed a lot of sensors that generate a massive amount of data and have a low computational power to process this data. So, the data needs to be offloaded to the cloud edge, which is more demanding from the network viewpoint [10]. In this case, when URLLC meets high broadband, the sub-6 GHz band may not withstand it due to the high interference and capacity limitations.

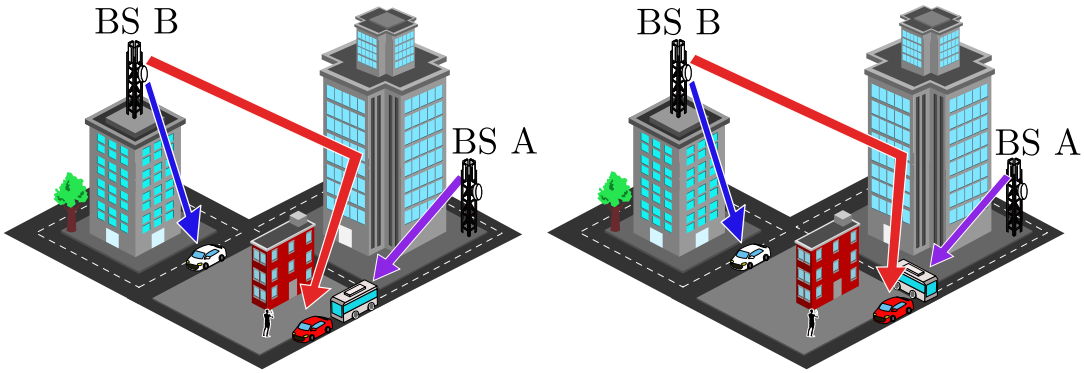
The Millimeter Waves (mmWave) band emerged as the enabler of future multi-gigabit communications. The Third Generation Partnership Project (3GPP) began to offer support for mmWave in 5G, through the Frequency Range (FR) 2, which is located at the mmWave band. The mmWave band spams from 30 to 300 GHz, but slightly lower frequencies are also classified as mmWave, for instance, 26 and 28

GHz. Notwithstanding the vast spectrum available, propagation at high frequencies is still an engineering challenge, with limited propagation range and severe Non-Line of Sight (NLoS) attenuation [11]. To compensate, high-gain directional antennas and dense deployment are required, leading to another drawback: mobility restrictions. Link blockage and antenna directionality, i.e., small aperture angles, demand beamforming for initial access, and beam tracking for mobility support. However, more beamforming also means more delay and overhead, due to the channel estimation before the beam alignment. Increasing the delay and overhead is the opposite of our aforementioned goal of minimizing latency and increasing reliability.

This research aims to optimize the 5G New Radio (NR) handover procedure, which is one of the factors that might enable mobility in URLLC mmWave 5G NR networks. The 5G NR baseline handover, originally proposed in Release 15 [12], is similar to the Fourth Generation of Mobile Communication (4G) handover [13] and is not seamless for low-latency applications, since there is a period of total data disconnection during handover execution, that can last for dozens of milliseconds and impede mobility for applications such as autonomous vehicles [10] and smart ambulances [14]. The 5G NR conditional handover mechanism proposed by the 3GPP in Release 16 [1] reduces these shortcomings but still has a margin to be improved [15]. In this thesis, we optimize handover for high-mobility and low-latency applications, such as Vehicle to Everything (V2X), based on the hypothesis that the well-informed handover decisions may result in better performance than traditional handover techniques. So, we focus on vehicular applications, such as autonomous vehicles, that need to constantly communicate with the network Base Stations (BSs) spread along the streets and roads.

To illustrate this, consider the situation in Figure 1.1. The red car finds itself in an uncertain handover situation, depending on the white bus' route. In a greedy handover technique, as soon as the Line of Sight (LoS) with the BS A is blocked, the handover to BS B is executed even if BS B offers only an NLoS channel because BS A's LoS is totally blocked. The handover to BS B is the better option when the white bus makes the turn, as Figure 1.1a shows, because the blockage situation will persist. However, if the white bus goes straight, as shown in Figure 1.1b, the better option is to endure a small blockage period than risk failing the handover and losing the connection with BS A completely. Making handover decisions like these is not trivial and is only possible due to situational awareness or prediction of future Channel State Information (CSI) .

Our approach is to formulate the handover as an optimization problem, constrained by the capacity and URLLC requirements. This optimization model has a thorough knowledge of the CSI providing an upper bound to assess other handover mechanisms, such as the 3GPP 5G NR baseline and conditional handover methods.



(a) (b)

Figure 1.1: Vehicular mobility scenario with mmWave BSs and blocking obstacles. In (a), the white bus takes the same route as the red car, blocking the red car's line-of-sight with BS A for a long period. Thus, the red car is handed over to BS B, which provides a non-line-of-sight link. In (b), on the occasion that the bus turns at the corner, the handover is not necessary and the red car has line-of-sight for most of the time.

The results obtained from this optimization if compared to the 5G NR handover methods show that an improvement in handover-related metrics, such as a reduction of 68% in the ping-pong effect rate. We can conclude from our assessment that the 5G NR handover methods, the baseline and the conditional one, can be improved to make better handover decisions and mitigate unnecessary delays, thus reducing delays and handover failures.

Though it is not feasible in real-time applications, the optimization leverages the use future predicted information in the handover decision process. Acquiring a partial knowledge of the future CSI is feasible and supported by recent advancements in machine learning, image processing and sensors embedding. Therefore, we propose a dynamic programming heuristic based on blockage prediction that approximates the optimization model performance. In our evaluation, the heuristic presented 10% and 7% less time under outage than the baseline and conditional handover methods, respectively.

1.1 Scope

This thesis focuses on handover optimization to achieve URLLC-compliant performance in the context of the 5G NR mmWave band in a blockage and mobility scenario. Communications at the mmWave band are heavily affected by the blockage caused by obstacles, such as the human body, trees, cars, etc. The 5G NR baseline handover method is only based on signal strength, making it more susceptible to handover upon blockage episodes [16]. As illustrated in Figure 1.2, the Target BS, which receives the mobile user via handover, must surplus by a margin the received

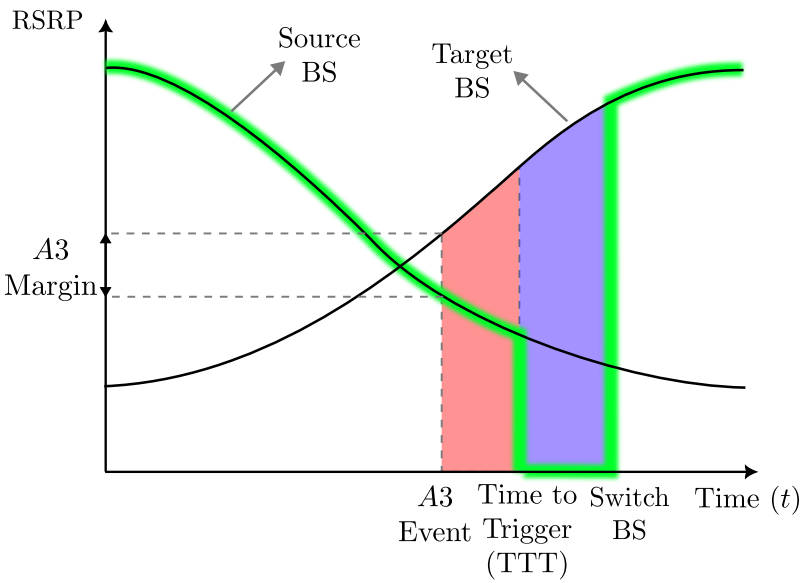


Figure 1.2: This figure depicts the baseline handover process. Whenever the Source BS's received power plus a margin is lower than the Target BS's received power the handover event is triggered and the Target BS is monitored for the Time To Trigger to ensure this condition of surplus holds.

power of the Source BS, which is the BS currently hosting the mobile user. A handover is only triggered if this condition holds for a period, called Time to Trigger (TTT). If the margin and TTT are not set adequately to the user velocity and the channel condition found in the scenario, unnecessary handovers might be triggered by blockage episodes. Restricting the number of handovers arbitrarily to avoid frequent connection breaks is a naive approach that might not exploit the full network potential, even in severe blockage conditions, resulting in a capacity decrease. Additionally, handover failures and ping-pong handovers are more likely to happen in high-handover rate scenarios. The former causes prolonged data connection loss for several seconds, while the latter causes congestion and increases overhead due to excessive control messages and context exchange between BSs. In scenarios with high blocking obstacle density and elevated interference, the network is more vulnerable to such effects and URLLC requirements are unattainable.

The 3GPP proposed the 5G NR conditional handover to mitigate baseline handover propensity to link failures. In conditional handover, two events must happen before executing the handover. The first event prepares the BSs that fulfil the event condition by adding these BSs to the execution list. Only the BSs in the execution list are monitored in the next phase until one meets the second event condition. Once this second condition is met, the user realizes the handover to the eligible BS. Figure 1.3 depicts the conditional handover preparation and execution events, each one with a received power margin and TTT. The conditional handover reduces Radio Link Failures (RLF) and handover failure and improves mobility performance

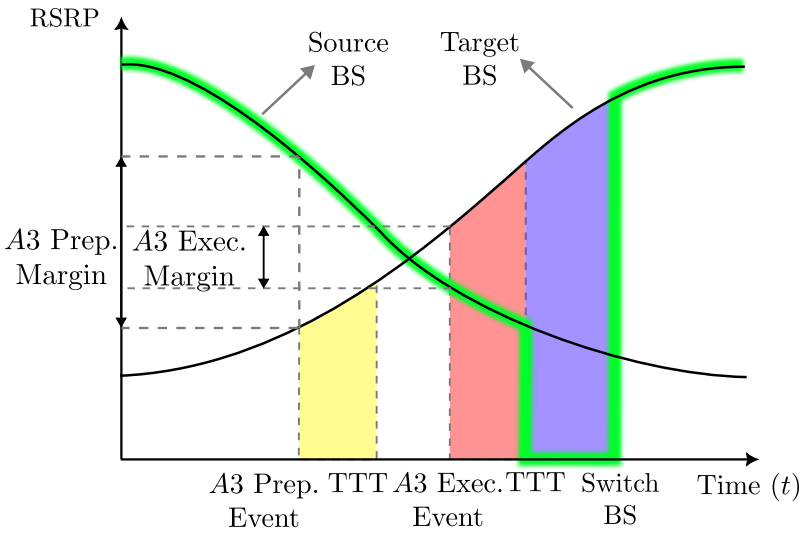


Figure 1.3: This figure depicts the conditional handover process. Different from the baseline handover, there are two events, the preparation and execution events, with different margins.

as the speed increases [15]. However, the overhead caused by unnecessary preparations and lengthy handover execution phases are some of the conditional handover handicaps.

Therefore, our primary goal is to find an optimal method to make handovers. Thus, we propose a Mixed-Integer Linear programming (MILP) problem that jointly optimizes the handover decision and performance indicators, like capacity and outage. Due to the knowledge of the whole CSI (for all BS and for all the time), the optimization can assess handover mechanisms with a performance upper bound and optimize wireless network site deployment for enhanced mobile user experience. Once we know the performance boundaries given by the optimization, we propose a heuristic handover method based on blockage predictions. The optimization is impractical due to the execution time and the necessity of all the channel data before problem-solving. However, it is proved in the literature that a few seconds ahead is predictable as long as the information from cameras or Light Detection and Ranging (LIDAR) sensors are available [17], for example. Finally, in our assessment, the proposed heuristic achieved a performance closer to optimal than the baseline and conditional handover due to its better handover decisions based on the predictions.

1.1.1 Motivation

The 5G NR raised new use cases for the mobile network supported by the enhanced capacity in contrast with the LTE. The URLLC use cases comprehend latency-sensitive applications such as factory automation, tactile Internet, and Intelligent Transportation System (ITS). Nowadays, URLLC is popularly known for the short

packets (32 - 200 bytes) sent with high reliability (10^{-5} of packet error probability) and low-latency (1 ms at layers 2/3) [8]. However, some applications might require a high throughput connection combined with the URLLC characteristics, such as autonomous vehicles.

The Vehicle to Infrastructure (V2I) use case motivates the combination of URLLC and Enhanced Mobile Broadband (EMBB). Due to the many sensors needed for safe transit, autonomous vehicles are envisioned to generate nearly 750 Mbps of data [18]. Such a huge amount of data will need to be offloaded to the cloud edge for processing, given that the vehicle-embedded hardware is not powerful enough to perform the computations, and for economic reasons, it can not be too expensive. Consequently, a fast-response, reliable, and broadband channel is fundamental for this system to be functional and safe.

Furthermore, the lack of performance boundaries restricts the analysis to the context of handover methods and scenarios tested. For this reason, we propose an optimization problem that jointly optimizes the handovers and performance indicators, providing a ground truth for comparing with other handover mechanisms. Investigating metrics like delivery rate, Shannon capacity and ping-pong handovers between the optimization and 3GPP proposed handover mechanisms it is noticeable the performance gap among them. Achieving such performance requires that the optimization gathers all the CSI beforehand, but despite the impossibility of having such information, it motivates the use of predictions to enhance the handover decision-making.

1.1.2 The problem

The mmWave bandwidth is the main enabler of a broadband channel, though it lacks reliability. With short-range links and sensitivity to line-of-sight blockage, the signals at mmWave bandwidth are prone to connection loss and handovers. Connection reestablishment and handovers may generate delay overhead for the user, as it probably interrupts data connection for several seconds.

Indeed, in the 5G NR, during the baseline and conditional handover procedures, while the User Equipment (UE) synchronizes with the new BS, the UE can not exchange data. This period without a data connection is called Handover Interruption Time (HIT) and may last up to 65 ms [19]. Besides, the handovers may fail due to the advance or delay in triggering the handover process. Considering the blockages, the handovers might be more frequent, aggravating the effect of HIT and handover failures on the UE experience.

The results in this thesis demonstrate that the 5G NR baseline handover is prone to a high handover rate and handover failure in a vehicular mobility scenario.

Because of the greedy nature of the baseline handover to search the BS with instantaneous highest received power, the handover decisions lead to handover failures and poor channel base stations, ultimately resulting in increased outage intervals. The conditional handover partially mitigates these negative effects but still takes handover decisions that compromise the network performance. In certain configurations, the conditional handover prepares an excessive number of base stations and delays the handover executions, bestowing the user with the decreasing channel quality of the source base station as the user moves farther away. Hence, it is critical to the network and user performance that the handover decisions, i.e., when and to which base station handover, lead to good and long-lasting channel conditions. Attempting to improve handover decision-making, we develop both an optimization model for establishing performance boundaries and a blockage prediction-aware heuristic handover method to make better handover decisions.

1.1.3 Related Works

Some plausible solutions to the above problem are discussed here. First, the naive approach is to reduce the number of handovers, be more cautious, and stick with the same BS for longer. However, depending on the UE velocity and trajectory, the same BS cannot offer an adequate channel for the UE needs during the connection interval. The higher the UE velocity, the faster the handover must happen to avoid handover failure and to provide a better channel by associating the UE with the new Target BS. One way to overcome this is tailoring the handover parameters to the user situation to force the handover to trigger when leaving the BS area and not when the channel fluctuates or a short blockage episode happens [20].

The second solution is to duplicate the UE incoming data on more than one BS to speed up the process of context exchange and reduce the HIT [21, 22]. This method can overload the network channels and interfaces between the BS or between the BS and the Core Network (CN). Finally, the UE can maintain a connection to more than one BS simultaneously, having redundancy when one of the BS goes out of range. This approach leads to BS overload and increases the overhead between the UE and the BS and between the BSs. These strategies provide redundancy for the UE data connection, however, the handovers are at least doubled, considering a primary and a secondary BS. In some cases, the UE power constraints might restrict the possibility of keeping association with two BSs, although the secondary connection is not active the whole time.

The 3GPP proposed the 5G handover in the earlier releases heavily inspired by the LTE handover [13]. Nonetheless, the baseline handover has limitations in the high mobility scenario, such as high mobility [23] and mmWave scenario [24]. In

these scenarios, the baseline handover is prone to handover failures and excessive ping-pong handovers, increasing the network load and the accumulated user interruption time. Hence, improvement alternatives were proposed such as predictive handover [25] or dual connectivity handover [4], for instance. These works increase the handover completion success rate and downlink transmission reliability, although such approaches could also lead to excessive control plane overhead and latency.

Further, the conditional handover was proposed to overcome the baseline high failure rate and time under outage in mobility scenarios [26]. The conditional handover has particular issues related to the preparation and execution phases [15]. These issues can be mitigated with custom parameter tuning and the aid of machine learning, as in [27], which increased the conditional handover success rate.

Optimization approaches are common in wireless networks, like in resource allocation studies, but handover optimization is not trivial due to the handover process dynamics. For example, in [28], the authors penalize the handover occurrences in the objective functions, resulting in handover inhibition and a lower network throughput. Instead of acting as a handover mechanism, the optimization though provides an upper bound for the other handover mechanisms. Similar to [29] we use a heuristic to optimize the utility function and achieve a near-optimal performance. This method is possible due to the possibility of predicting blockage [30] or received power [31] and by doing so increase the handover mechanism awareness of the CSI to come.

1.1.4 Proposed Solution

In this thesis, we propose to efficiently make the handovers, by making better decisions in the long term. Despite the HIT associated with the handover, when necessary, the handovers benefit the UE. Differentiating between channel fluctuations and the UE moving closer to/farther from a BS, and knowing the duration of a blockage episode are important to define if the blockage is bearable or if a handover should be triggered. Thus, we define an optimization model that knows this information before the handover decisions. This model takes the handover decisions to optimize a performance indicator, avoiding handovers that would degrade the performance. The model's purpose is to establish an upper bound to a proposed heuristic handover mechanism, the 5G baseline and conditional handover methods. The heuristic handover takes advantage of a short prediction of the blockage status to assess if a handover pays off. Using dynamic programming the heuristic handover method explores the handover possibilities in an interval by evaluating a utility function. Hence, the heuristic approach takes advantage of the optimization model's concept without needing the whole channel information, but with near-optimal performance.

1.1.5 Methodology

The methodology adopted in this thesis is to establish the boundaries of performance and assess the efficacy of handover techniques compared to these boundaries. Discrete event simulation is the main tool used in our evaluation. The scenario of interest is a vehicle in a straight route moving with constant speed while travelling through a path with obstacles spread around and BSs alongside. The obstacles are capable of blocking the line-of-sight between the UE and the BSs, which increases the path loss attenuation the mmWave signal bears. We have implemented the 3GPP 5G NR baseline and conditional handover respecting the standard timings, power levels, and signalling specifications. Thus, the simulation is written in Python covering the user movement, blockage status check to calculate the CSI, and the 3GPP 5G NR basic functions and handover protocols.

A MILP optimization problem provides the theoretical upper bound. We modelled the MILP problem using the Gurobi solver's Python library. The model is fed with an input instance which contains the CSI for all simulated timeslots for each pair BS-UE and every possible handover opportunity. During the solution process, the handover possibilities are added to the solution according to the handover and the other constraints, such as capacity and delay thresholds. Some objective functions can be plugged into the model depending on the metric to be optimized. An example of an objective function is to maximize the sum of capacity for all the timeslots.

Afterwards, we propose a handover technique based on dynamic programming to be compared with the baseline and conditional 3GPP handover techniques. All of them, including the optimization, are evaluated with the same input instances, though the optimization needs a preprocessing first to generate the CSI while the others calculate the CSI in each simulation step. Finally, we can obtain metrics like capacity, and packet delivery rate and examine handover rate, ping-pong handover ratio, and handover failures. This set of metrics allows us to investigate the strengths and weaknesses of each technique, how close they are to the optimal, and the key points to developing a handover algorithm.

1.2 Outline

This thesis provides a background on mmWave and 5G NR, which are the fundamental concepts for the technical chapters ahead. Then we dive into the optimization model, detailing the constraints, and discussing results and lessons learned. Finally, we conclude this thesis by summarizing the contributions and planning our future works.

1.2.1 Chapter 2 - Background

In this chapter, we introduce the key aspects of 5G NR and mmWave channel characteristics. First, the mmWave channel characteristics are described. Also, we describe how to model a mmWave link on system-level simulation. Second, we bring the reader into contact with the 5G NR main features, and the details of the baseline and the conditional handover processes.

1.2.2 Chapter 3 - Joint Handover-URLLC Optimization Model

In Chapter 3, we describe the system and optimization models adopted. The optimization model jointly optimizes the handover decisions and performance metrics with a customizable objective function. We present three objective function possibilities that reflect different operation modes of handover mechanisms: capacity maximization, capacity maximization and delay assurance, and capacity maximization and outage minimization.

Moreover, we also present the dynamic programming handover based on blockage predictions. The heuristic handover is supported by the evidence that the optimization exploits the future CSI to make the handover decisions wisely, but is not feasible to assume all the CSI information beforehand. Thus we detail the scenario that provides such predictions and how these predictions are used to provide insightful data to the handover heuristic.

1.2.3 Chapter 4 - Results

In Chapter 4, our goal is to assess the handover algorithms previously described, the 3GPP 5G NR baseline, conditional and our proposed heuristic comparing with the proposed optimization upper-bound. First, we profile our simulation environment and the parameters used in the evaluation. The evaluation is divided into two parts, one dedicated to the impact of the handover configuration parameters, and the other regarding the blockage density and velocity. The results show that the proposed heuristic handover outperforms the baseline and conditional handovers in many aspects, proving its improved efficiency in the handover decision-making process, but still has a performance to the optimization.

1.2.4 Chapter 5 - Conclusions

In this chapter, we draw our conclusions and plan our future work based on the outcomes of this research. As complementary solutions to the optimization model, we intend to develop a Machine Learning (ML) framework.

1.3 Contribution

In this Section, we summarize our technical contributions. The benchmark in the literature for handover mechanisms are the 3GPP 5G NR baseline and conditional handover mechanism, which we detail in Chapter 2. In Chapter 3, we introduce the MILP optimization model and assess the model complexity and performance and a dynamic programming heuristic handover based on blockage predictions. Hence, we compare the 5G handover mechanisms with the proposed heuristic handover and the optimization ground truth in Chapter 4. Our contributions are:

- An MILP joint optimization model for handover and performance indicators, such as capacity maximization and outage minimization. This model establishes an upper bound for future investigations on handover.
- Multiple objective functions for the MILP optimization model, maximizing capacity, joint capacity maximization and delay minimization, and joint capacity and reliability maximization.
- We proposed a heuristic handover technique that uses limited time blockage predictions to make the handover decisions. A dynamic programming algorithm periodically finds the best handover opportunities leading to a performance closer to the optimal than the baseline and the conditional handover.
- A 5G NR baseline and conditional handover implementation on a discrete event simulation. We reproduced the handover protocols' timing, control messages, and measurement checks to approximate the implementations of the standardized handover methods. All the handover methods described in this thesis are evaluated in a mobility environment with mmWave BSs and blocking obstacles randomly scattered that can interfere with the channel status.
- An assessment of the handover configuration parameters influencing the handover mechanisms. We assessed the Time to Trigger (TTT) and received power margin, parameters used to decide whether a handover will be triggered in the 5G NR baseline and conditional handover techniques.
- An analysis of the velocity and blockage density impact. By varying these two parameters we find interesting outcomes, for example, a non-monotonic trend in handover rate when the blockage density increases.
- The results showed that the gap between the baseline and optimization is large in most metrics tested and that the conditional and the heuristic improve the baseline performance. The proposed heuristic handover is better than the conditional due to its knowledge of the incoming CSI, for example by

reducing outage time, packet loss and delayed packets. The evaluation of the handover mechanisms against the upper bound provided by the optimization gave valuable insights into the shortcomings and points of improvement to focus on when designing a handover technique.

1.4 Dissemination

This section lists the papers written during the D.Sc. project:

- da Silva Brilhante, Davi, José F. de Rezende, and Nicola Marchetti. "Handover optimisation for high-capacity low-latency 5G NR mmWave communication." Ad Hoc Networks 153 (2024): 103328.

Chapter 2

Background

In this Chapter, we will present the mmWave channel and the 5G features, focusing on handover and mobility management. The goal of this chapter is to provide the reader with the tools to understand the further chapters of this thesis. In Section 2.1, we describe the mmWave channel characteristics, opportunities, and challenges. Section 2.2 describes the motivation and goals of 5G, as well as its Physical Layer (PHY) layer capabilities and handover methods.

2.1 The millimeter Wave Channel

The mmWave band figures between the enabling technologies of current and future generations of mobile wireless communication. Such an expectation is justified by its most valuable asset: a massive almost non-occupied spectrum. From 30 GHz to 300 GHz, where the mmWave band is located, had been considered idle until recently [32]. This characteristic is very promising because it helps with at least two problems of sub-6 GHz networks, high co-channel interference and lack of free spectrum, whether licensed or not. Whereas sub-6 GHz technologies, such as LTE, are limited in data rate, mmWave networks allow multi-gigabit data rate [33].

Despite the potential of the mmWave, it presents certain drawbacks. As the signal frequency increases, the free-space attenuation also increases, decreasing the signal propagation range [34]. To cope with this handicap, highly directional antennas compensate for the propagation loss. Conveniently, the small wavelength of mmWave signals permits the use of large antenna arrays to obtain high directionality, which leads to low interference among different devices. With directional antennas, another challenge arises, transmitter and receiver antennas must be aligned for maximum gain [11]. Therefore, steerable antennas and alignment algorithms are required. In addition, the signal in the mmWave band has low material penetration and low refraction capabilities, that is, the signal barely steers around objects and is likely to be completely blocked. Attenuation through materials reaches values such as 10

to 20 dB through a car [35] depending on the receiver position and ranges from 10 to 32 dB through the human body [36]. This attenuation level can completely prevent the signal from reaching the receiver. In short, the high attenuation and low penetration of mmWave signals require using directional antennas, and numerous blockages make the mmWave link unstable.

Owing to the small wavelength in the mmWave band, it is possible to significantly reduce the antenna size and use a printed circuit antenna array. Advanced circuitry and antenna design allow the antenna array to be electronically steered adjusting each antenna gain and phase shift, a technique called beamforming. There are three beamforming architectures: analogue, digital, and hybrid. Analog beamforming is cheap and power-efficient; however, it lacks precision and supports only one data stream. Digital beamforming supports multiple data streams and is more accurate, however, it is also expensive and consumes more power due to the demanding processing of the baseband information. Hybrid beamforming provides a trade-off of the analogue and digital advantages and disadvantages to achieve an efficient yet simple solution [37].

Depending on the antenna and beamforming design parameters, like the number of elements and the space between elements of an antenna array, the radiation patterns of the beams may change. A bank with presets of phase shifts for each antenna element is called a beam codebook. Beamforming with a beam codebook involves choosing the preset that maximizes a certain metric after probing several preset beam codewords, in a process called beam sweeping. Consequently, if the number of beam codewords increases, the beam sweeping overhead also increases.

With beam sweeping, the transmitter and the receiver can exploit the full capacity of the antenna arrays. The naive beam sweeping algorithm is an exhaustive search in which the transmitter and receiver try every combination of beams that the beamforming provides. This algorithm is cumbersome and may be prohibitive for low-latency communications, particularly when the beams are narrow or the devices are moving. Opportunities arise for developing better beam sweeping algorithms, as diverse approaches can be found in the literature [38].

Figure 2.1 depicts a codebook-based beam sweeping and beam selection procedures. After the beam is established, one starts the beam management, which monitors the state of the connection in order to provide beam maintenance when necessary. Mobility intensifies the urge for beam management, and as the speed of the mobile user increases, the harder is to maintain the directional connection, until the point when a handover is necessary. Regarding the handover, the beams must be redefined, due to the change in the mobile user position and the beam misalignment that it causes. Handover strongly affects the performance and user experience. In mmWave communications, high mobility requires efficient beamforming techniques

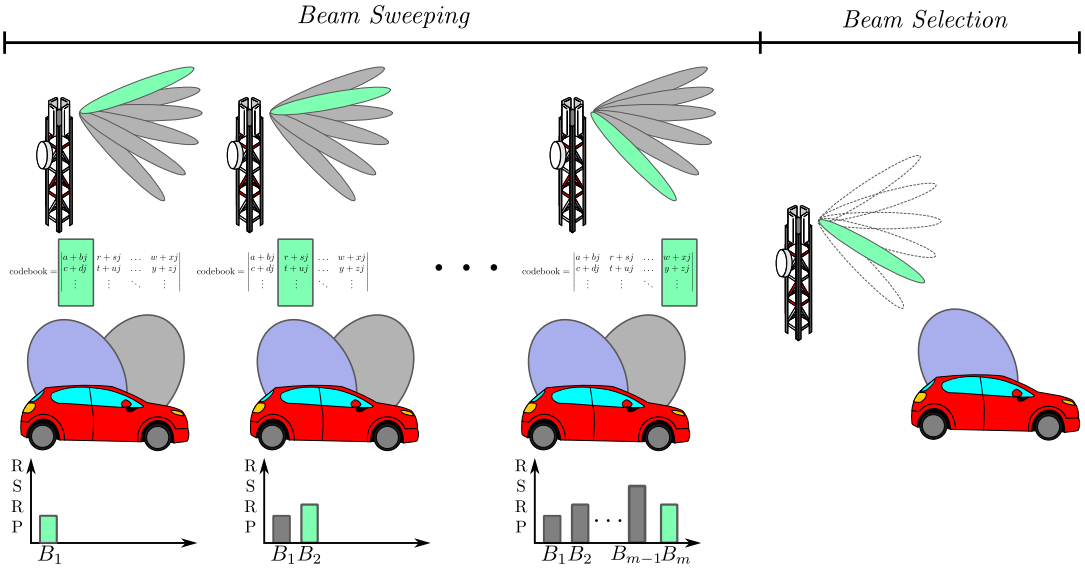


Figure 2.1: Codebook-based beam sweeping and beam selection procedures

and reliable handover mechanisms for assuring the best achievable performance [39].

Exploiting mmWave for high capacity, low latency, and ultra-reliability is possible, though the challenging trade-off among these Key Performance Indicators (KPI) requires each involved mechanism to be sharply designed [40]. In this thesis, we shed light on the handover procedure, which in the mmWave context involves the proper detection of the handover necessity, choosing the right BS to handover to, and the efficient beam alignment with the new BS [33]. For instance, a blockage episode may trigger a handover from BS A to BS B, and then, when the blockage is over, another handover back to BS A, generating overhead and UE power consumption. This kind of situation severely harms the performance of mmWave links and should be avoided at all costs. In the remainder of this Chapter, we will present the 5G NR and its handover procedure, to further introduce our contribution in Chapter 3.

2.2 The fifth-Generation of Mobile Wireless Network

In the fourth quarter of 2017, the 3GPP launched the first release of 5G, i.e., release 15, marking the start of 5G. Since then, the 3GPP launched three more releases, and the 5G deployment started around the globe. In April 2022, 5G networks were launched in more than $1,947$ cities across 72 countries [41]. According to Ericsson, the coverage forecast for 2028 is to reach more than 80% of the world population[42].

5G NR makes a significant technological leap from Fourth Generation of Mobile Communication (4G) LTE. In numbers, 5G NR increased the data rate to over 1 Gbps at 28 GHz bandwidth [43], reduced round trip time latency to nearly 10

ms [44], provided coverage enhancement [45], and improved overall network energy consumption by 70% [46]. Technologically speaking, the 5G NR relies on state-of-art enabling technologies, such as multiple service frequency bands, Beam Division Multiple Access (BDMA), and flexible numerology [47, 48]. By numerology, we mean the Orthogonal Frequency Division Multiplexing (OFDM) Subcarrier Spacing (SCS), the presence of cyclic prefix, and the OFDM symbol duration. Such a set of capabilities makes the 5G NR a versatile and powerful mobile wireless network standard. Indeed, the 5G NR is the first technology that can offer support to the three following use cases:

- **Enhanced Mobile Broadband (EMBB):** The goal of EMBB is to deliver near-wired connection features. This use case aims at high throughput connectivity, which has more and more a popular appeal and calls the attention of most users, mainly driven by the widespread of Augmented Reality (AR) and Virtual Reality (VR), video streaming services, videoconferencing, and enhanced gaming.

- **Ultra-Reliable and Low-Latency Communication (URLLC):** This use case covers delay sensitive, below 10 ms, and high reliability, 10^{-5} packet loss probability. The applications under the URLLC umbrella are, for instance, industrial automation, Vehicle to Vehicle (V2V) communication, and Intelligent Transportation System (ITS).

- **Massive Machine Type Communication (MMTC):** the MMTC component provides low-power and wide-area connection for Internet of Things (IoT) devices. Many setups of latency and throughput are available in MMTC, due to the variety of sensors applied in IoT applications, for example, real-time sensors, such as emergency detection sensors, and delay tolerant sensors, like in agricultural networks. We highlight smart cities, smart health care, and V2I as MMTC applications.

Such use cases are not exclusive, rather some applications are mixing two use cases, widening the range of network requisites. A flexible PHY layer and state-of-the-art technologies enable such a broad range of applications in 5G. In the next section, we define some aspects of the PHY layer related to the handover and mobility which were modelled in our simulations.

2.2.1 PHY Layer

One of the 5G main characteristics is the flexibility it offers to providers and users. For example, 5G supports two frequency ranges, sub-6 GHz, called FR1, and a

η	Subcarrier Spacing	OFDM Symbol Duration	Cyclic Prefix	Frequency Range
0	15 KHz	1/14 ms	Normal	FR1
1	30 KHz	1/28 ms	Normal	FR1
2	60 KHz	1/56 ms	Normal	FR1, FR2
2	60 KHz	1/56 ms	Extended	FR1, FR2
3	120 KHz	1/112 ms	Normal	FR2
4	240 KHz	1/224 ms	Normal	FR2

Table 2.1: 5G supported transmission numerologies [1, 2]

mmWave bandwidth, called FR2. The FR1 includes 410 MHz up to 7.1 GHz channels, which are the most used ones for the deployment of 5G networks and the same used on LTE and other Wireless Local Area Network (WLAN) technologies. Due to their low and mid-range frequencies, and popular use in recent years, these frequencies are easy to work with except that they are already used extensively, increasing co-channel interference and limiting the maximum amount of contiguous channels. The FR2 is located in the mmWave bandwidth, starting at 24.2 GHz up to 52.6 GHz, offering large portions of unutilized spectrum. However, using FR2 is challenging due to the high free-space attenuation the mmWave suffers from, as detailed in Section 2.1.

With Carrier Aggregation (CA), multiple bandwidths can be used, in a totally transparent fashion to the user. Additionally, the provider can set multiple numerologies in the same bandwidth. The frame and subframe have fixed duration, 10 and 1 ms, respectively. The number of slots varies according to the numerology, but the number of OFDM symbols in a slot is always 14. Hence, depending on the chosen SCS, defined by the parameter η, the number of slots, and by consequence, the number of OFDM symbols changes. The numerologies are detailed in Table 2.1.

Another important part of PHY in the 5G is the cell search and selection procedure, which also synchronizes time and frequency. The synchronization is periodic in the downlink and uplink and the periodicity can have some different values. In the downlink, the measurements are transmitted in the Synchronization Signal (SS) bursts, with period $T_{SS} \in \{5, 10, 20, 40, 80, 160\}$ ms. Each SS burst is composed of several SS blocks. An SS block is a group of 4 OFDM symbols that carry the Primary Synchronization Signal (PSS), Secondary Synchronization Signal (SSS), and the DeModulation Reference Signal (DMRS), where the latter can be used for Reference Signal Received Power (RSRP) measurements in the SS block. In the uplink, the synchronization occurs in the Random-Access Channel (RACH), with period $T_{RACH} \in \{5, 10, 20, 40, 80, 160\}$ ms. The RACH procedure is triggered when UE needs to make a handover, at the initial access procedure, and after a beam/radio link failure.

In the FR2, the synchronization is also responsible for the beam selection. The

Event	Triggering Condition	Purpose
A1	SBS RSRP becomes better than a threshold	Intra-RAT
A2	SBS RSRP becomes worse than a threshold	
A3	Neighbor BS RSRP becomes offset better than SBS RSRP	
A4	Neighbor BS RSRP becomes better than a threshold	
A5	SBS RSRP becomes worse than a threshold T_1 and the neighbor BS RSRP becomes better than a threshold T_2	
A6	Neighbor BS RSRP becomes offset better than a secondary BS RSRP	
B1	Inter-RAT neighbor BS RSRP becomes better than a threshold	Inter-RAT
B2	SBS RSRP becomes worse than a threshold T_1 and the inter-RAT neighbor BS RSRP becomes better than a threshold T_2	

Table 2.2: 3GPP 5G NR handover events [3].

mmWave BS may send one SS block of an SS burst per beam, while the UE stays fixed at the same beam. The process of sending one probe on every beam is called beam sweeping. At the end of the SS burst, the UE may change its beam and receive a freshly started SS burst. The same will occur at the RACH, but from the UE to the BS, sending a random access preamble transmission MSG1 (Message 1) and receiving a random access response MSG2. In the FR2, there can be a maximum of 64 SS blocks per burst, and the SS bursts can span through 5 ms, i.e., half frame. This exhaustive approach probes every pair of beams between UE and BS and is considered a naive and cumbersome beam selection method.

2.2.2 Handover

The 5G NR handover process is similar to the 4G LTE handover. In general, the UE gathers RSRP measurements of the surrounding BSs and reports them to the BS it is associated with. If at least a pair of BS fall into some specific conditions defined by the handover event, then the handover is triggered. We summarize the 5G NR handover events in Table 2.2. The handover events are classified into intra-Radio Access Technology (RAT), when the handover occurs between two 5G BS, for example, and inter-RAT, such as a 5G to LTE handover. From now on, we will call the BS to which the UE is associated by Serving Base Station (SBS) and the candidate BS for receiving the UE via handover by Target Base Station (TBS).

The Radio Resource Control (RRC) is a layer 3 control plane protocol created by 3GPP since the Universal Mobile Telecommunication System (UMTS). The primary role of RRC is to control and configure all the radio resources of lower layers

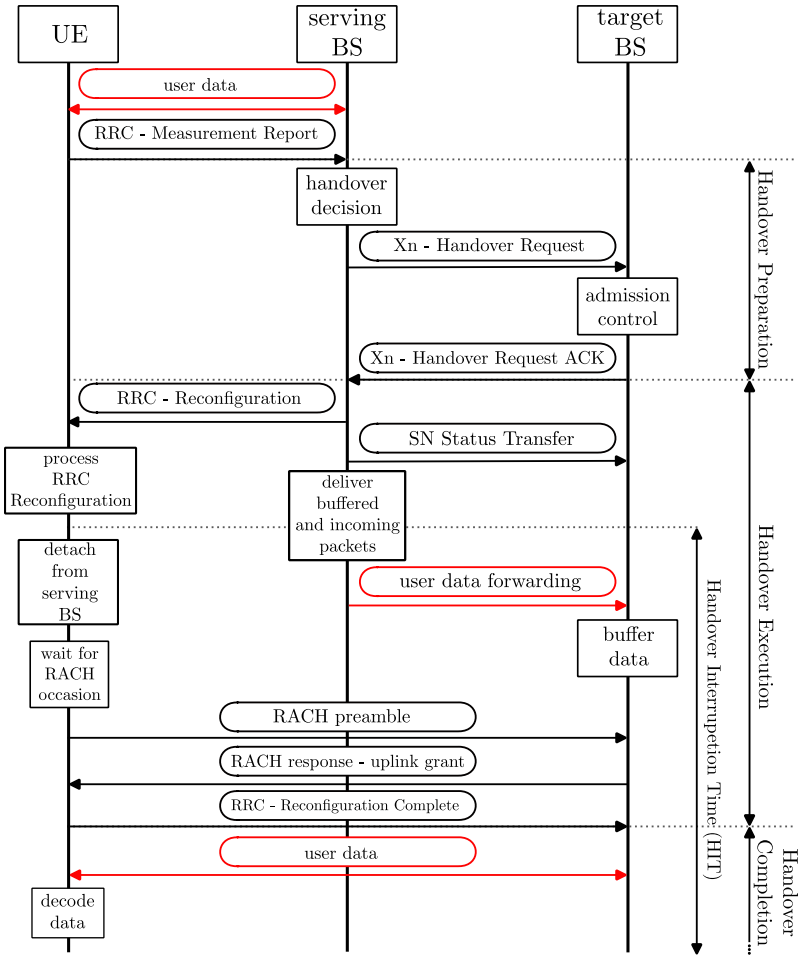


Figure 2.2: 3GPP 5G NR baseline handover protocol

to enable communication between UE and BS. Thus, the RRC protocol provides all the UEs with connection life-cycle, mobility management, Quality of Service (QoS) management, CA management, and security functions. The functions of interest related to the RRC are signal monitoring and mobility management, especially handover. The message exchange in the handover procedure is defined in the scope of the RRC protocol, as depicted in Figure 2.2.

The legacy RRC provides two states: *RRC_CONNECTED* and *RRC_IDLE*. The 5G NR introduced the *RRC_INACTIVE* mode. The *RRC_INACTIVE* is an intermediate mode between the *RRC_CONNECTED* mode and the *RRC_IDLE* mode, with energy-saving purpose. Different from the *RRC_IDLE*, the *RRC_INACTIVE* keeps a connection with the BS, including mobility management, but with suppressed signalling. The last serving BS receives from the CN all the data and control messages to the UE, which only receives short control messages with downlink decoding details, such as modulation scheme and which slot to transmit. While inside the Radio Access Network (RAN) Notification Area (RNA), which is a list comprised of a single or multiple BS, the UE can transition to *RRC_CONNECTED*, and any BS in the RNA can retrieve the data and signalling directly from

the last serving BS. Therefore, the UE is still connected from the CN point of view. Thus, the state transition delay from *RRC_INACTIVE* to *RRC_CONNECTED* when there is data to transmit is shorter than from *RRC_IDLE*, favouring bursty data transmissions. The mobility in the *RRC_INACTIVE* state differs from that mobility in the *RRC_CONNECTED* state. In the *RRC_INACTIVE*, if the UE moves inside the RNA, there is no need for a location update, reducing power and resource consumption. Otherwise, if the UE is outside its original RNA, it will need a cell reselection procedure. On the other hand, in the *RRC_CONNECTED* state, a network-controlled handover is necessary for connection maintenance during UE mobility.

RSRP Measurements

The UEs in *RRC_CONNECTED* mode periodically monitor the reference signals sent in the Synchronization Signal (SS) block bursts transmitted by every BS. Using the DeModulation Reference Signal (DMRS) the UE can measure the RSRP, which will further guide the handover decision. However, the raw signals, such as DMRS, are inadequate for the handover decision process due to the channel's fast fluctuations [49]. Two filters are applied to the raw measurements to alleviate the channel's fluctuations [22]. The first filter operates at Layer 1 or PHY Layer and averages the last N_{L1} raw RSRP samples P_{raw}. The output of the Layer 1 filter for every t_{L1} Layer 1 measurement sample is given by Equation (2.1).

$$P_{L1}(t_{L1}) = \frac{1}{N_{L1}} \sum_{i=0}^{N_{L1}-1} P_{\text{raw}}(t_{L1} - i) \tag{2.1}$$

The output of the Layer 1 filter is then filtered again by the Layer 3 or the Radio Resource Control Layer filer. This filter receives a measurement samples every ω_{L3}. The output of the Layer 3 filter for every t_{L3} Layer 3 measurement sample follows the Equation (2.2). Figure 2.3 depicts the scheme of RSRP filtering as proposed by 3GPP.

$$P_{L3}(t_{L3}) = \alpha P_{L1}(|P_{L1}|) + (1 - \alpha) P_{L3}(t_{L3} - 1) \tag{2.2}$$

Where $P_{L1}(|P_{L1}|)$ is the last element output by the Layer 1 filter, i.e., the most recent Layer 1 filtered measurement sample. The Layer 3 filter is an Infinite Impulse Response (IIR) or an Exponential Weighted Moving Average (EWMA) filter with a forgetting factor α [50]. The definition of the forgetting factor α is related to the filter coefficient k

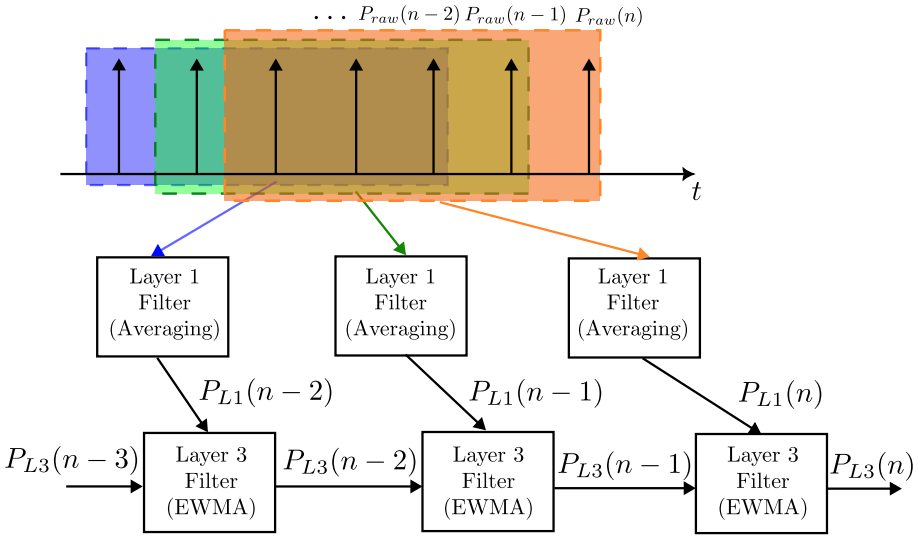


Figure 2.3: RSRP Layer 1 and Layer 3 Filters

$$\alpha = \left(\frac{1}{2}\right)^{\frac{k}{4}} \tag{2.3}$$

The Layer 3 filter mitigates fast channel fluctuations that could generate erroneous channel quality measurements by smoothing the Layer 1 filter output samples with previous Layer 3 filter samples. The Layer 3 measurements define the UE operation regarding the handover, handover failures, and synchronization assessment, and poor measurements would cause severe instabilities and performance jeopardizing.

Handover Protocol

In general, the handovers happen when the UE is in *RRC_ CONNECTED* state and a measurement falls into one of the conditions defined by the handover events, as summarized in Table 2.2. The event A3, for example, is triggered when a neighbour BS RSRP gets better than the SBS RSRP by an offset plus a hysteresis margin. Then, this neighbour BS becomes a TBS and the UE starts a timer called Time to Trigger (TTT). During the TTT, the TBS and SBS RSRP are monitored. The measurements are reported to the SBS in the form of Message Report (MR). If the handover triggering conditions hold, then the handover will proceed. Otherwise, the handover is interrupted with an MR of leaving triggering condition.

The TTT is useful to prevent the network from handover the UE in case of a fast channel fluctuation. Thus, the TTT reduces the ping-pong effect and the handover failures. Also, the offset and hysteresis margins contribute to avoiding unnecessary handovers. The event A3 triggering condition can be defined as

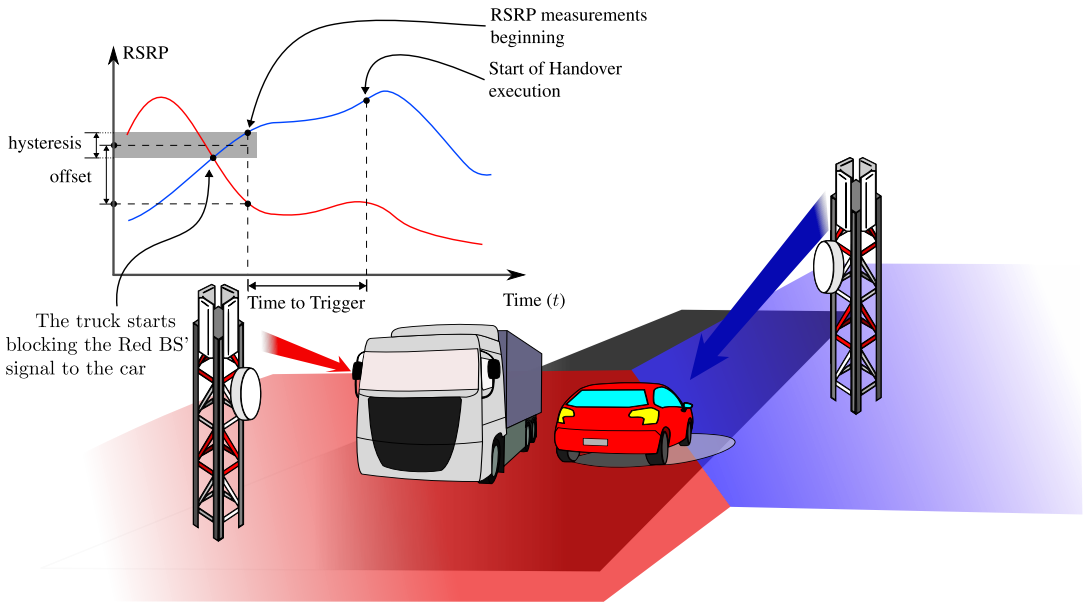


Figure 2.4: 3GPP 5G NR handover scheme. In the figure, the truck blocks the mmWave link between the car and the red BS, triggering a handover to the blue BS.

$$R_T > R_S + O + H, \tag{2.4}$$

wherein R_T and R_S are the target and serving BS's RSRP, respectively; O and H are respectively, the offset and hysteresis handover parameters, expressed in dB. Fig. 2.4 shows an RSRP variation that triggers a handover; due to the blockage caused by the truck, the serving BS's RSRP (red curve) drops till the point it is exceeded by the target BS's RSRP (blue curve) and after the TTT, if the triggering condition held, then the handover is executed.

Additionally, the TTT strongly influences the handover triggering when the UE is moving. If the velocity increases, the TTT must decrease and vice-versa. As a high-speed UE reaches the border of the SBS in a shorter period of time, the TTT must be short enough to respond to such change. Otherwise, the UE will reach the SBS edge without having initiated a handover to a TBS. The opposite is also true, e.g., at a lower speed, the UE is more susceptible to a fast-fading induced RSRP variation until it reaches the BS border, so the TTT must be set large enough to avoid handover when signal fluctuations affect the RSRP. In this case, if the TTT is too short, the number of handovers will unnecessarily increase, also increasing the overhead and the Handover Interruption Time (HIT), worsening the UE performance.

When the TTT finishes, if the handover triggering condition is true, the SBS will make a handover decision. The decision is made by the SBS based on the MR and the Radio Resource Management (RRM) information which sends a handover

request on the interface *Xn* to the chosen TBS. Before proceeding with the handover, the TBS performs admission control, assuring that resources can be granted to the incoming UE. Then, a handover request acknowledgment is sent to the SBS, which notifies the UE with a handover command message preparing the UE for detaching from the former SBS and attaching it to the new one.

Once the UE detaches from the SBS, there is no data connection established until it fully synchronizes with the new BS. The synchronization must be established in the downlink and the uplink, using the SS burst and RACH periods. This period while the UE remains disconnected is called HIT and if it is too long, it can make delay-sensitive applications impractical [14, 51]. In severe conditions, like NLoS propagation, high network load, and poor cell coverage, the total experienced latency grows [52]. It is important to reduce or eliminate sources of delay, such as HIT and re-transmission delay, to enable URLLC and capacity-demanding applications.

Ping-pong and Handover Failures

Besides the HIT, other factors might prevent a good handover performance. The Time of Staying (ToS) in a BS can be used to assess the quality of a handover. The ping-pong effect is when the UE goes back and forth between at least two BS in less than a Minimal Time of Staying (MToS). In other words, a ping-pong happens when the ToS of a BS is shorter than the MToS. If this happens, the intermediate handover is considered unnecessary, only generating overhead, network congestion, and power consumption. Ping-pong-caused handovers are expected to increase in the 5G, due to network ultra-densification and narrow beamwidth at FR2 [53]. Reducing the ping-pong handover rate is a critical goal to pursue on the hunt for enabling URLLC and improving KPI.

Likewise, handover failures may be caused by wrong handover decisions and result in a connection re-establishment. A handover may fail due to a delay in changing to a new BS or changing too soon to a new BS. These handover failures are frequently caused by a Radio Link Failure (RLF). A radio link failure happens when the SINR of a BS measured at the UE is inferior to the out-of-synchronization (Q_{out}) threshold. When more than N310 consecutive SINR samples are measured in this condition a T310 timer is started. At the end of the T310 timer, we consider that a radio link failure happened. However, if at least N311 consecutive samples exceed the in-synchronization (Q_{in}) threshold before the T310 expires, the latter is interrupted and the radio link does not fail. A radio link failure during the handover will lead the handover to failure, for instance, if the UE does not receive the *RRC Reconfiguration* message from the SBS (too late handover) or if the TBS can not receive the *RRC Reconfiguration Complete* from the UE (too early handover).

Another possible way of handover failure is when the *RRC Reconfiguration* procedure or the RACH with the TBS cannot be completed. An RACH is not completed if, in a limited time, the UE sends the RACH preambles and not receive a RACH response from the BS. When the SBS sends the *RRC Reconfiguration* message to the UE, a T304 timer is initiated. The timer is stopped when the RACH with the TBS is completed. Otherwise, if the timer expires before the RACH is complete, the handover is considered to have happened too early and a failure occurs.

RRC Connection Re-establishment

When a handover fails or an RLF happens the RRC Connection Re-establishment procedure must be initiated. There are several reasons to trigger a re-establishment procedure, which can be summarized in the following five reasons [12]:

1. when a Radio Link Failure (RLF) happens,
2. when a handover fails,
3. when the maximum number of random access attempts is reached,
4. when the configured maximum number of Radio Link Control (RLC) retransmissions is exceeded,
5. when the UE cannot comply with the configuration set by *RRC Connection Reconfiguration* message.

RLF and handover failures were described previously. Reaching the maximum number of random access attempts is when the UE sends the RACH preambles but does not receive the RACH response from the BS for a maximum number of retries, even increasing the RACH power transmission in each new attempt [12]. The RLC is a layer 2 protocol, which is responsible for Automatic Repeat Request (ARQ), segmentation and error detection [54]. When the RLC is in acknowledged mode the UE waits for BS acknowledgement of the RLC packets and sets a polling timer. This timer consists of the product between the round-trip time and the maximum number of retries. If the RLC packets are not acknowledged before the polling timer expires, the UE considers it an RLF. In the last case, the UE can not meet the configurations specified by the BS receiving the UE.

Upon detecting one of the above conditions, the UE must stop the data connection with the SBS, suspending all the ongoing transmissions. Subsequently, the UE performs a cell selection process, in which the UE periodically scans for the available BSs nearby by jumping to frequencies that may contain an SSB, also called Channel Raster. A T311 timer starts at the beginning of the cell selection process and is stopped only when the UE finds a suitable cell [3].

A BS is selected if the UE can detect the SSB and successfully decode the PSS and SSS to get the Cell ID. Henceforward, the UE decodes the Master Information Block (MIB) which contains information about the numerology, such as the subcarrier spacing, and the UE must locate the System Information Block message 1 (SIB1), which contains information regarding the resources availability and scheduling. Then, the BS and UE advance to the RACH procedure and the UE can start the connection re-establishment mechanism after finishing RACH [1].

The re-establishment mechanism initiates with the UE sending a connection re-establishment request to a BS and waiting for the acceptance. The acceptance is sent only after the BS finds the UE context in the network, and in case the connection re-establishment is with the SBS only a context modification is needed. The UE context contains information about the UE, such as its radio and security capabilities, current Internet Protocol (IP) address and QoS requirements. Otherwise, the new BS should make a context setup procedure to accommodate the new UE. Afterwards, the UE must acknowledge that the re-establishment is completed and the BS does the context modification procedure. The process is completed with an exchange of *RRC Reconfiguration - RRC Reconfiguration Complete* messages [55].

In the case of an RLF, for instance, after the T310 expires and the RLF is declared, a T311 is triggered [3]. While this timer counts, the UE must reselect and send the connection re-establishment request to the prior SBS. Once the request is sent, the T301 timer is initiated to wait for the BS connection re-establishment response. This example of connection re-establishment when an RLF happens is depicted in Figure 2.5.

Besides the connection re-establishment protocol, the whole re-establishment process includes the time to detect the loss of connection, the scan and selection of a new BS, the time to acquire the system information, and the time to perform RACH [3]. This process may take several seconds [56] and the UE can not transmit any data, degrading the network performance and resulting in a poor user experience. Thus, the root cause of the connection re-establishment must be tackled to reduce its occurrences, and consequently, minimize the UE connection interruptions due to the procedure's duration.

Therefore, how can one diminish the handover handicaps and make the handover process more efficient? Without modifying the 3GPP default procedure, the only way is to carefully tune up the network parameters involved in the handover to optimize the desired KPI. However, with external aid or adaptations to the handover procedure, like camera images [57] or dual connection [21], the network performance can be significantly enhanced for achieving URLLC requirements.

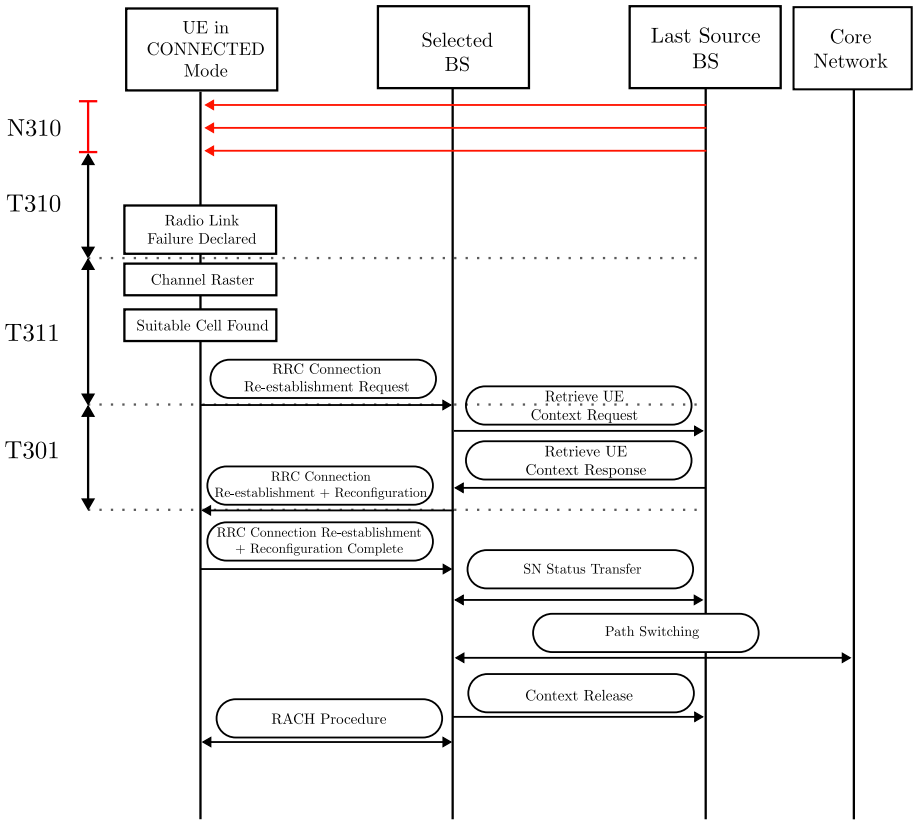


Figure 2.5: RRC Connection re-establishment procedure in case of an RLF

2.2.3 5G NR Conditional Handover

Aiming to improve mobility performance and reduce RLF, the 3GPP proposed conditional handover in release 16. The conditional handover divides the handover process into two phases: preparation and execution [58]. Figure 2.6 depicts the 3GPP conditional handover protocol. A Target BS assessment criterion is defined for each phase, usually offset margin of RSRP or RSRQ. In the preparation phase, the neighbour BS for which the RSRP is greater than the preparation threshold is included in the execution list. The BSs in the execution list are monitored until one triggers the execution condition and becomes the TBS. In this mechanism, when the preparation phase finishes, the UE does not need to exchange handover control data with the SBS, but can still exchange data. Assuming that the UE is distancing from the SBS as the handover process occurs, eliminating sooner the dependency on the SBS reduces the probability that a handover will fail due to Radio Link Failure with the SBS. In the execution phase, the UE only needs to communicate with the TBS for uplink synchronization.

In the preparation phase any handover events summarized in Table 2.2 can be used as the preparation condition. When a Neighbour BS (NBS) meets the preparation condition this BS is requested for a conditional handover. After admission control, if the request is accepted, this BS enters the execution list. The execution

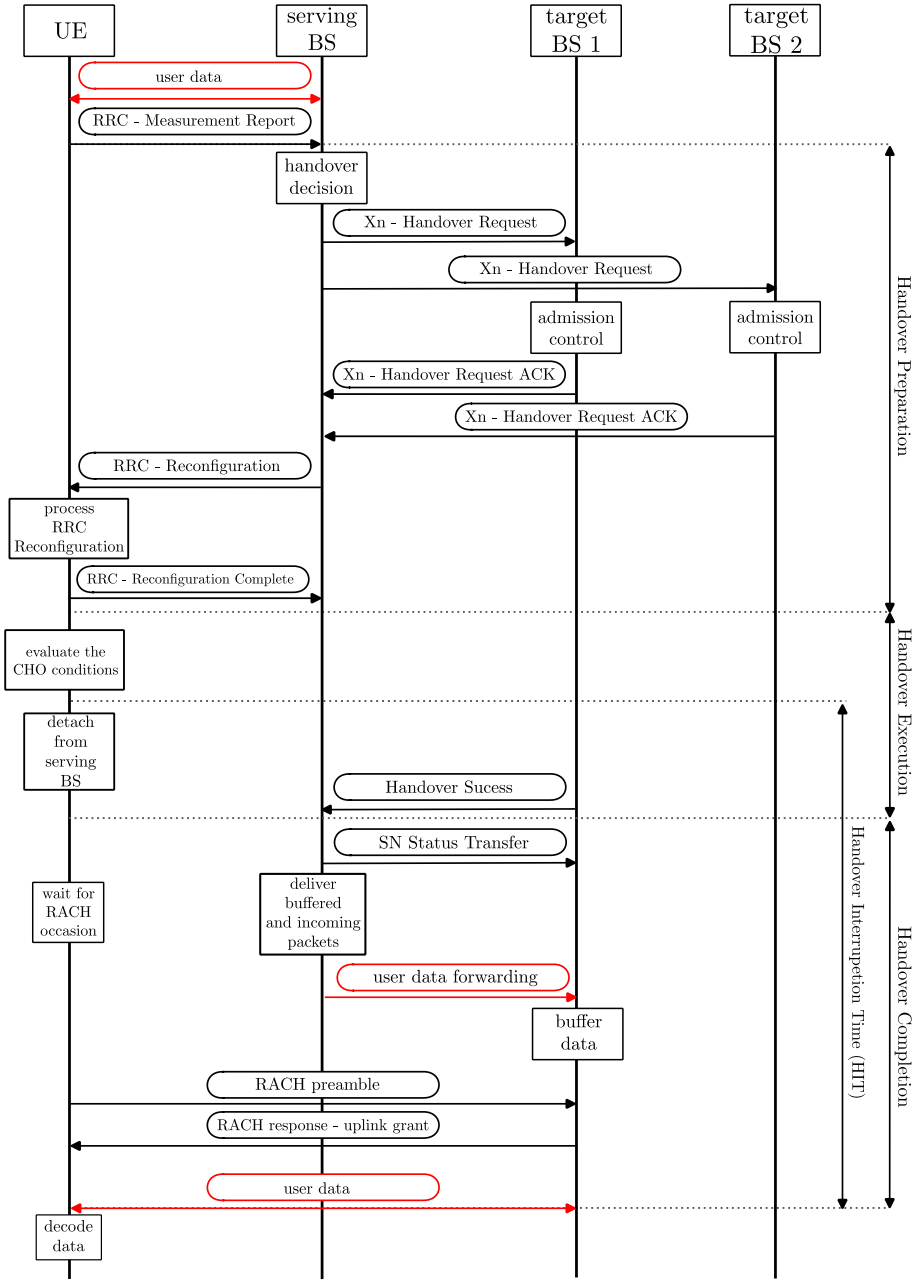


Figure 2.6: 3GPP 5G NR conditional handover protocol

list is a limited list of BSs that will be monitored for the execution condition. Once the execution list is formed, it is sent to the UE with the other data necessary to proceed with the handover and the UE begins to monitor the BSs for the execution condition [15]. Although the preparation phase allows the mitigation of RLF due to earlier TBS preparation, if not correctly configured it can result in inflated signalling overhead due to an excessive number of prepared BSs [15, 59].

On the other hand, in the execution phase only the events A3 and A5 can be used as the execution condition, with RSRP or RSRQ [26]. Until any BS in the execution list meets the execution condition, the UE can exchange data with the SBS. However, there is no handover control message to be exchanged with the SBS, reducing the handover failure probability. Data may still be lost if the signal of SBS is too poor and none of the prepared BS has fulfilled the execution condition yet. Whenever a prepared BS meets the execution condition the UE starts to reconfigure itself to perform the RACH procedure for uplink synchronization. The UE is associated with a new SBS if the RACH is successfully finished.

A BS might be released from the execution list if the RSRP falls below the release threshold [58]. This threshold for releasing a BS is usually smaller than the preparation and execution thresholds. Also, the preparation list length is limited to a certain number of BSs, and BSs may not be accepted if the list is already full. Thus, removing the BS in the worst condition may open a spot in the list for a better BS, benefiting the handover and leading to a higher success probability.

While in the execution phase, the UE may be able to send data packets to the SBS until one BS' RSRP falls in the execution condition. Once the execution condition is met, the UE reconfigures its radio to tune up with the TBS and do the uplink synchronization with the RACH. Though it is reduced, the HIT problem persists [22], because the UE stays without data connection until fully synchronized with the target BS.

The conditional handover is not fail-proof. Two conditions may result in a handover failure: a radio link failure during the early stages [59], or an unsuccessful RACH [26]. In the former, when there is a radio link failure between the SBS and the UE, the handover control messages cannot be delivered, so the handover does not proceed. In the latter, a timer called T304 is triggered when the RACH with TBS starts, if this timer ends before the RACH is complete the handover also fails due to incomplete synchronization with the TBS.

Nonetheless, the conditional handover procedure has a better recovery plan in case of handover failure. Instead of doing the *RRC Connection Re-establishment* process, the UE can take advantage of the prepared BSs, if available [26]. When the conditional handover fails the UE follows the connection re-establishment steps up to the cell selection. If the selected BS is in the prepared list the UE may access

this BS using the configuration information stored during the preparation phase and proceed with a normal handover to haste the connection recovery. Otherwise, the UE must continue with the default connection re-establishment process.

2.3 Related Works

Throughout this whole chapter, we described the basis of the 3GPP 5G NR, regarding the numerology, signalling and initial access. Besides, we focused on the two handover techniques specified in the standard, the baseline and the conditional handovers. The baseline has many similarities to the LTE handover [13] and is non-suitable for high mobility in mmWave scenarios. Handover failures and excessive number of handovers increase the number of periods under outage. This negatively impacts the user quality of service, undermining applications of interest, such as URLLC mobility and massive sensor data transmission.

For a minimal change of the 3GPP 5G NR handover procedure, some authors propose the fine-tuning of the handover parameters, such as the TTT and the offset margin [60, 61]. Farooq et al. [61] developed a Linear Programming (LP) optimization model which outputs the optimal A3 and A5 handover parameters for maximizing the downlink RSRP, the handover success rate and the the network load balance distribution between frequency bands. This LP is compared to ML algorithms, outperforming all algorithms, with the eXtreme Gradient Boosting (XGBoost) achieving the lowest Root-Mean-Square Error (RMSE).

A well-known way to increase the handover robustness and reduce data interruption delay is dual connection. The 3GPP added support for dual connection since LTE Release 12 [62]. In dual connection schemes, the UE keeps a connection with two BS, adding diversity and redundancy. In normal circumstances, the two BS will simultaneously serve the UE. If the channel with one BS is under poor propagation conditions, the link with the other BS will supply the communication. In the case of handovers, dual connection increases the handover success rate and decreases the rate of failed downloads [4, 21]. However, dual connection schemes in an urban dense scenario considerably increase the number of handovers, which added to the extra signalling for maintaining two BS, may overburden the network with overhead [63, 64].

To tackle the weaknesses of the baseline handover, the 3GPP proposed the conditional handover, previously introduced in this chapter. The goal of 3GPP with conditional handover is to increase handover robustness and, secondarily, to reduce the HIT by keeping a connection with the SBS until the last moment possible. Despite its proven efficacy in reducing RLFs and handover failures for medium and high speeds [26] compared to the 5G baseline handover, the conditional handover

has performance flaws. Some shortcomings of conditional handover include overhead caused by early preparations and unnecessary resource allocation on the target BS, which occurs when the execution conditions are met too late or this target BS is not selected for handover [15]. Besides that, in high interference scenarios or aggressive fading, if a BS cannot hold the preparation condition, the execution does not go forward. So, the handover is delayed and more likely to fail [59].

Side information can be used to trigger conditional handover. In [65], the authors successfully mitigated the throughput degradation when walking humans block the link LoS while assuming that cameras would be available and active. However, the UE power consumption and the user's privacy due to having an active camera are not considered. In the same way, the pre-connection handover may exploit multiple antennas simultaneously, where an antenna subset is used for data transmission and the remaining antennas are used for pre-connection [66], instead of using all the antennas for data connection.

The RACH period may be the main factor behind the HIT having an extended duration, so suppressing RACH might be a viable strategy to reduce HIT. Attempting to make a seamless handover, the RACH period assumes a time-synchronized network so that the UE synchronization can be skipped, configuring a RACH-less handover. In [67], the authors evaluate the impact of contention-free RACH associated with a decision tree algorithm to penalize the beams of the BSs prepared that led to handover failure. In [68], the authors define a new event that triggers the acquisition of the timing advance. The timing advance is crucial to eliminate the synchronization with the TBS and enable the RACH-less handover. The RACH-less handover effectively reduces the outage and HIT, though with the cost of increasing signalling overhead.

In this thesis, we propose a Mixed Linear Integer Programming (MILP) optimization problem which optimizes the handover and KPIs related to the URLLC mobility applications. The literature on wireless network optimization is rich in resource allocation works, that consider several network features, such as interference mitigation, delay minimization, and aggregated throughput maximization, for example. In [69], authors investigated the power and resource allocation of a single-cell URLLC V2V network. The optimization model is formulated as a transmit power optimization, aiming for interference reduction, but subject to the probabilistic length of a data packets queue. The latency and reliability are improved, though one considers only LoS between the vehicles. Likewise, in [70] the optimization objective function maximizes the network utility subject to the queueing delay, but these metrics, namely allocated power, delay, and data rate, are de-coupled and optimized separately.

A bi-objective MILP approach to handover is introduced in [28]. One objec-

tive function maximizes a utility function while the other penalizes the handover occurrences. The handover penalization objective function attempts to serve the maximum number of UEs with the minimum number of handovers. The constraints assure the minimum utility value and that the resources offered to the BS will not exceed the maximum available amount. The outcome of this model is a handover shortage, evidenced by the lower network throughput compared with the Max Signal-to-Noise Ratio (SNR) algorithm, where the UE associated with the BS that offers the best downlink channel condition indicated by the instantaneous SNR.

In [29], the objective function formulation involves three metrics: the expected rate, the expected number of UE with demand not fulfilled, and the expected number of handovers. Each one of these metrics has an individual weight in the objective function for increasing or decreasing the optimization emphasis on a specific metric. So, it is possible to tailor the performance, for example, augmenting the focus of the optimization on the expected rate when capacity-demanding applications are predominant. Moreover, the MILP is turned into a reinforcement learning model, solved by two methods Markov Decision Process (MDP) and Deep Deterministic Policy Gradient (DDPG). The results show an increase in the probability of realizing less than 10 handovers and a decrease in the number of outage events.

Although there are numerous works on resource allocation and URLLC optimization, there is a lack of works that model the handover process. When a handover modelling is found, it does not consider the handover limitations entirely or just penalizes the network for doing handovers. Additionally, the handover optimization is frequently solved by other means, such as relaxations or ML. We consider important the establishment of an upper bound, so that other methods, such as heuristics and ML, are evaluated by how far they are from the optimum. Also, many optimization models approach the handover too simplistically, compromising the optimization results application to real scenarios.

In the case of the optimization proposed in this thesis, it is not practical to assume that it will be implemented for making real-time handover decisions. First, it is not reasonable to assume that the UE or BS will have knowledge of all the further blockage events for a relatively long period ahead. Second, the time to compute the model solution is considerably longer than the period available to make a handover decision in a URLLC system. However, the optimization model is a powerful framework for handover optimization evaluation. With the bound the model provides, we can assess how far a handover optimization strategy is concerning the optimization model. Thus we approached the handover with a heuristic handover technique.

Due to the low computational complexity, heuristics are also explored as a sub-optimal solution for complex analytical models. In [71], a bi-objective optimization

problem is used to minimize the effects of blockage and ensure load balancing. Because of the complexity involved in the problem solution, a heuristic is proposed to solve the problem in polynomial time and a semi-distributed fashion, offloading computations to Cloud RAN (CRAN). Regarding the link maintenance, in [72], a MILP problem optimally picks which link maintenance method is better: handover, beamforming, or Device-to-Device (D2D) relay. A heuristic then approximates the optimization problem solution with lower complexity.

Machine learning has been popularized in wireless networking problem-solving and is also used in conditional handover optimization. To make the conditional handover more reliable and reduce the overhead, in [73] the authors use Deep Learning for trajectory prediction. Predicting the UE path allows the preparation of the BSs alongside the user trajectory, limiting the number of preparations and releases, consequently, reducing overhead. In [27], the conditional handover preparation success rate significantly increases to almost 98% and the resource reservation time is reduced. For this, a conditional handover preparation event A2 occurs and triggers a measurement report, then BS runs a Deep Neural Network (DNN) to predict the execution event based on an RSRP time series.

With the evolution of image processing and ML, prediction in wireless networks has been successfully applied. In [31], with the support of images, the authors developed an ML prediction model for received power that achieved high accuracy. The proposed mechanism predicted the received power 500 ms ahead with 3.4 dB of RMSE. But images are not the only input data that work to this end, sub-6 GHz [74] channel data and Light Detection and Ranging (LiDAR) [30, 75] were also employed in the literature. Knowing beforehand hundreds of milliseconds can be exploited for handover decision enhancement [10].

Moreover, ML algorithms can result in highly accurate solutions for numerous problems because of their generalization power. In terms of handover, reinforcement learning techniques are predominant in the literature. In reinforcement learning, a learning agent can adapt the actions taken to the changes in the environment, to maximize the received reward and establish an optimal policy. This approach is suitable for the dynamic situations experienced in wireless network problems. For example, some reinforcement learning algorithms successfully applied to handover optimization, are Q-learning [76], Multi-Armed Bandit (MAB) [16, 53, 77], multi-agent reinforcement learning [29, 78] and deep reinforcement learning [79].

We can also highlight other ML techniques, besides reinforcement learning. In [80], the target neighbour BS choice is based on classification algorithms, such as gradient boosting and random forest, and the RSRP data history. Neural Networks are also frequently employed in prediction, decision-making, and pattern recognition, for example in fingerprint technique [25]. The fingerprint technique consists

of a database segmented by geographical area, that when queried infers from the input the user's location and returns the best answer for this area. A Long-Short Term Memory (LSTM) neural network predicts the channel conditions for handover latency reduction, with previous Received Signal Strength Indicator (RSSI) measurements of several BS given as input. Then, after the input is successfully labeled, the BS which is predicted to offer the best Shannon capacity is chosen as TBS to receive the UE.

2.4 Conclusion

This chapter details the mmWave propagation characteristics, the 3GPP 5G NR standard and its handover methods, the baseline and the conditional handover. The baseline handover has the convenience of comparing RSRP levels between the Source BS and the Neighbouring BSs, which also has a low level of complexity. However, the greediness of the algorithm may lead to handover decisions that are local maximums, making handover to the first BS that meets the handover condition, for instance. The immediate and short-sighted handover decisions may result in excessive handover failures, radio link failures, and poor channel conditions. The conditional handover attempts to cope with these faults by defining two conditions with different offsets. Because the preparation offset is usually lower than the execution offset, the conditional handover can capture the trend of the BS's RSRP and handover only if the trend is non-decreasing.

The 5G NR baseline and conditional handover methods can be improved with the aid of parameters fine-tuning and optimization, heuristics and ML. But what is the margin for improvement? In Chapter 3, we attempt to answer this by proposing a MILP optimization problem which optimizes handover by not restricting or minimizing the number of handovers but triggering the handovers that optimize an objective function, for example, maximizing the capacity sum. Nevertheless, the optimization is focused on being an evaluation and network planning tool rather than being a functional handover technique. There are two reasons for this, first, the time required to compute a handover decision is prohibitive, and second, the optimization needs all the CSI before the optimization starts. Thus, we also propose a heuristic handover technique that follows the optimization concept of having future information. Recent advancements in ML and device capabilities to embed sensors like cameras and LIDARs enable blockage and RSRP predictions for a few seconds ahead with high accuracy. The heuristic handover uses blockage predictions to find the handover opportunities that maximize a utility function using dynamic programming.

Chapter 3

Enhanced 5G NR URLLC Handover Mechanisms

In this chapter, we introduce the MILP optimization model. The optimization model is our first contribution, aiming to establish an upper bound for handover optimization algorithms. Different objective functions can be used, defining the optimization focus on capacity-only, joint capacity and delay, or joint capacity and reliability. Moreover, the constraints guarantee the handover conditions and general system behaviour, such as the UE can only be associated with one BS. Another contribution presented in this chapter is the handover heuristic which exploits short-time blockage predictions to make future-aware handover decisions using a dynamic programming algorithm.

3.1 Introduction

In the previous chapter, we discussed the potentials and challenges of using mmWave for high throughput communication, the features of 3GPP 5G NR and its handover procedure. One of our concluding remarks is that the 5G NR handover in the mmWave band is unable to provide full mobility support when URLLC and high throughput are jointly targeted. In dense BS deployment scenarios, the number of handovers increases due to the small cell coverage range that frequently leads the UE to cross a cell border. This situation worsens the delay and capacity performance because the handover-generated control messages will increase the delay and the interference elevation may reduce the data rate.

In addition, mmWave links are frequently interrupted by blocking obstacles, breaking network stability, and increasing the number of handovers. Frequent handovers increase overhead and add data interruption delay [16, 81], compromising the network performance. Although a naive solution might be to restrict the number of

handovers, they are necessary to keep the KPI at acceptable levels. The problem that arises is making the handovers at the right time and to the correct BS [57].

3.1.1 Contributions

To cope with the challenges mentioned above, in Section 3.2 we propose a MILP formulation that jointly optimizes different KPIs and the handover for vehicular mobility on a mmWave blockage scenario. We formulate the MILP framework to optimize capacity, delay and reliability. The framework offers three possible objective functions and their necessary constraints: (i) capacity maximization, (ii) joint capacity maximization and delay assurance, and (iii) joint capacity and reliability maximization. The optimal formulation provides performance boundaries for future research on handover, e.g., the optimal number of handovers when evaluating mechanisms based on heuristics. The model can also be used offline for network planning and adjustment of handover decision parameters.

Based on the optimization, we introduce in Section 3.3 a simplified version of the optimization model, focusing only on the handover constraints and the utility function. We propose a heuristic for handover based on blockage predictions to solve this sub-problem efficiently. The MILP optimization depends on having the CSI at hand in the moment of problem-solving and the computational time it takes to execute and solve the whole problem, which is too large considering the time window required to make a handover decision. From a simplified version of the problem, we propose a dynamic programming algorithm that is computationally simpler than the proposed MILP formulation. Instead of using perfect CSI, the heuristic handover is based on imperfect and time-limited blockage predictions. The dynamic programming algorithm and the shorter predictions provide satisfactory handover decisions with a lower computational load and feasible time for real applications.

To help the reader throughout this chapter, we provide the symbol table 3.1 which contains the symbols used in this chapter in the optimization and heuristic, in order of appearance.

3.2 Mixed Linear Integer Programming Model

To model our optimization problem, we use an observation period of T timeslots. Such limitation keeps the number of model constraints bounded, assuring the problem feasibility and allowing us to concentrate entirely on the lifecycle of the UE on the network, i.e., when the UE enters the simulation area and when it leaves. The problem is represented by one binary decision variable, $x^m_{n,t}$. The variable $x^m_{n,t}$ is equal to 1 if the UE n is associated with the BS m at time slot t, given that the

Table 3.1: Chapter 3 symbol table

Symbol	Description	Section
T	Simulation observation period	3.2
$x_{n,t}^{m}$	User equipment association decision variable for the user n with the BS m at timeslot t	3.2
$\mathcal{S}_{n,t}^{m}$	SNR between UE n and BS m at a given timeslot t	3.2
$\mathcal{P}_{n,t}^{m}$	UE n received power in dB from BS m at timeslot t	3.2
$\sigma^2_{\text{noise},dB}$	Additive white Gaussian noise power	3.2
$P_{tx,dB}^{m}$	BS m transmission power	3.2
$G_{tx,dB}^{m}$	BS m transmitter antenna gain in dB	3.2
$G_{rx,dB}^{n}$	UE n receiver antenna gain in dB	3.2
$P_{L,dB}(d_{n,t}^{m})$	Path Loss in dB between UE n and BS m at timeslot t	3.2
$d_{n,t}^{m}$	Distance between BS m and UE n at the timeslot t	3.2
$\mathcal{N}$	Gaussian random variable	3.2
μ	Gaussian random variable mean	3.2
σ	Gaussian random variable standard deviation	3.2
$\mathcal{C}_n$	UE n Capacity requirement	3.2
$\mathcal{R}_m$	BS m physical resource blocks	3.2
A_k^n	UE n packet k arrival time	3.2
τ	Time to trigger	3.2
$u_{n,t}^{p}$	Handover opportunity binary indicator variable	3.2
$\mathcal{C}_{n,t}^{m}$	UE n capacity with BS m in timeslot t	3.2.1
SC_η	subcarrier spacing for the chosen numerology η	3.2.1
η	5G NR chosen PHY Layer numerology	3.2.1
$\mathcal{Q}$	Set of handover candidate BSs	3.2.1
$y_{n,t}^{m}$	Message delay decision variable regarding UE n and BS m if there is a message to be sent in timeslot t	3.2.2
$\mathcal{D}_n$	UE n delay tolerance	3.2.2
K_n	Number of UE n data messages	3.2.2
$\mathcal{W}_k$	k-th message delay window	3.2.2
$\mathcal{I}_{n,t}$	UE n delay penalization function at timeslot t	3.2.2
Q_{out}	SINR Quality-out threshold	3.2.2
$v_{n,t}^{m}$	SINR below out-of-synchronization threshold Q_{out} indicator variable	3.2.2
κ	Out-of-synchronization penalization factor	3.2.2
$\mathcal{L}$	Blockage prediction window length	3.3.1
R_{Th}	Heuristic handover BS selection distance range threshold	3.3.1
$\mathcal{B}_{\text{pred}}(m,n,t)$	Blockage prediction for BS m and the UE n at timeslot t	3.3.3
$\mathcal{M}'$	Subset of candidate BSs selected by the heuristic handover algorithm	3.3.3
ell	Dynamic programming algorithm timestep	3.3.3
$\mathcal{C}(k,m,n)$	Shannon capacity calculated for the link between the BS $m \in \mathcal{M}'$ and the UE n using the predicted blockage status at k	3.3.3
p_{ho}	Heuristic handover penalization	3.3.3
$w_{n,t}^{m}$	Heuristic handover decision variable	3.3.3

BS has the minimal resources the UE needs to maintain its services online, and 0 otherwise.

Additionally, the SNR (or the SINR, alternatively) of the link between UE n and BS m, $\mathcal{S}^m_{n,t}$, is assessed at every timeslot t. Evaluating the BS SNR indicates whether the BS still can keep the UE application and also serves to calculate the UE attained capacity at each timeslot, which will further be used for assessing the optimization model. The SNR, in dB, is defined by Equation (3.1).

$$\begin{aligned}\mathcal{S}^m_{n,t} &= \mathcal{P}^m_{n,t} - \sigma^2_{\text{noise},dB} \\ &= P^m_{tx,dB} + G^m_{tx,dB} + G^n_{rx,dB} - P_{L,dB}(d^m_{n,t}) - \sigma^2_{\text{noise},dB}\end{aligned} \tag{3.1}$$

wherein, $\mathcal{P}^m_{n,t}$ is the UE received power; $\sigma^2_{\text{noise},dB}$ is the noise power; $P^m_{tx,dB}$ is the BS m transmission power; $G^m_{tx,dB}$ and $G^n_{rx,dB}$ are the BS m and UE n antenna directivity gain, respectively; and $P_{L,dB}(d^m_{n,t})$ is the path loss attenuation, as defined in Equation (3.2)[82], for the distance $d^m_{n,t}$ between BS m and UE n at the timeslot t.

$$P_{L,dB}(d^m_{n,t}) = \begin{cases} 61.4 + 20\log_{10}(d^m_{n,t}) + \mathcal{N}(\mu = 0, \sigma = 5.8), \text{ if LoS} \\ 72.0 + 29.2\log_{10}(d^m_{n,t}) + \mathcal{N}(\mu = 0, \sigma = 8.7), \text{ if NLoS} \end{cases} \tag{3.2}$$

Wherein, $\mathcal{N}(\mu = 0, \sigma)$ represents a zero mean Gaussian random variable with standard deviation σ. It is trivial to see that the path loss model sampling results in higher path loss for NLoS links with higher probability than for LoS links, given the same distance d, due to the greater σ and the other two model constants.

Furthermore, we consider that every UE has a capacity requirement $\mathcal{C}_n$. In the pre-processing phase, having the BS positioning, the UE and obstacles coordinates, the UE x and y velocity components, the Reference Signal Received Power (RSRP) and SNR are all calculated for each timeslot following the path loss model in Equation (3.2). Then, the optimization model receives as input the SNR per timeslot, $\mathcal{S}^m_{n,t}$, the BSs resources, $\mathcal{R}_m$, the user capacity constraint, C_n, and the user message arrival times, A^n_k. All these parameters are constants for the optimization model, only the binary decision variables $x^m_{n,t}$ are accounted for in the model.

Proceeding with a handover, as defined by 3GPP in the 5G NR standard, requires two conditions to be met. Considering an A3 event (refer to Table 2.2), the first condition is that the RSRP of the target BS must be greater than the serving BS's RSRP by a margin. Secondly, the former condition must hold for the TTT interval (henceforth denoted as τ). As the input instance has the RSRP for all

timeslots beforehand, it is possible to enumerate for each pair of BS all the handover opportunities according to the aforementioned handover conditions. The A3 event condition and the requirement it holds for at least a time equal to the TTT can be expressed as in Equation (3.3).

$$C_1: \ \mathcal{P}^q_{n,i} > \mathcal{P}^p_{n,i} + H + O \ \ \forall \ i \ \in [t, t+\tau] \tag{3.3}$$

However, a handover only happens from a serving BS p to a target BS q if the UE is formerly associated with BS p. The association is in turn defined during the problem-solving, as it is related to the decision variable $x^m_{n,t}$. Then, for each handover opportunity from BS p to whichever candidate BS, a binary indicator variable $u^p_{n,t}$ is added to the problem. Consequently, $u^p_{n,t}$ is only added to the optimization model if the handover opportunity exists and meets the condition C_1 in Equation (3.3). The variable $u^p_{n,t}$ indicates if the UE n was associated with BS p for a period equal to or greater than τ before the handover execution at t, that is a mandatory condition for a handover to happen.

$$u^p_{n,t} = 1 - min\left(1, \tau - \sum_{j=t-\tau-1}^{t-1} x^p_{n,j}\right) \tag{3.4}$$

In Eq. (3.4), $u^p_{n,t} = 1$ if the UE n is associated to BS p for the last τ time slots at least ($x^p_{n,j} = 1, \ \forall \ j \in [t-\tau-1, t-1]$). Otherwise, if the UE is not associated with BS p during the given interval, then $u^p_{n,t} = 0$. Observing if $x^p_{n,j} = 1$ in the interval $[t-\tau-1, t-1]$ reflects the measurement reports during the TTT mentioned in Section 2.2. Afterwards, the optimization decides whether a handover will happen based on the variables x and u. Next, we present the optimization model formulation in Eq. (3.5).

$$\mathcal{P}1: \max_{x,y} \quad f(m,n,t) \tag{3.5a}$$

$$\text{s.t.} \quad x^m_{n,t} \leq \left\lfloor \frac{\mathcal{R}_m \mathcal{S}^m_{n,t}}{C_n} \right\rfloor, \tag{3.5b}$$

$$u^p_{n,t} = 1 \rightarrow \sum_{q \in \mathcal{Q}} x^q_{n,t} \leq 1, \tag{3.5c}$$

$$u^p_{n,t} = 0 \rightarrow x^p_{n,t} = x^p_{n,t-1}, \tag{3.5d}$$

$$\sum_{m \in M} x^m_{n,t} \leq 1, \ \forall \ n \in N, \ \forall \ t < T, \tag{3.5e}$$

$$
\begin{aligned}
&\text{(3.5b)}, \forall\ n \in N,\ \forall\ m \in M,\ \forall\ t < T,\\
&\text{(3.5c)}, \text{(3.5d)},\ \forall\ n \in N,\ \forall\ p \in M,\ \forall\ t < T,\\
&x, y, u \in \{0, 1\}
\end{aligned}
$$

One may argue that knowing all T timeslots SNR beforehand is an impractical assumption for real-time applications. However, the optimization model constitutes a useful framework for offline applications, such as site planning and network parameter choice for handover enhancement. Besides that, the optimization problem sets boundaries for assessing the performance indicators in handover mechanisms, like the minimal number of handovers.

3.2.1 Constraints

The constraint (3.5b) assures that the UE is only associated with BS that can provide enough PRB to attain the capacity $\mathcal{C}_n$. The constant $\mathcal{C}^m_{n,t}$ represents Shannon's theoretical capacity, wherein $\mathcal{R}_m$ is the BS m bandwidth measured in PRB, and $\mathcal{S}^m_{n,t}$ is the UE SNR at timeslot t. In order to simplify notation, we omit in constraint (3.5b) the logarithm of the SNR and the PRB-to-bandwidth conversion, but it is shown in Equation (3.6), where SC_η is the subcarrier spacing for the chosen numerology η.

$$
\mathcal{C}^m_{n,t} = \mathcal{R}_m(SC_\eta \times 12)\log_2(1 + \mathcal{S}^m_{n,t}) \tag{3.6}
$$

Constraints (3.5c) and (3.5d) determine whether a handover will take place or not. The set $\mathcal{Q}$ contains all the target BS candidates to receive a UE handed over from serving BS p. As an indicator constraint, if the left-hand side condition is true, the right-hand side should happen. On the contrary, if the left-hand side condition is not true, then the right-hand side may be false or not. Hence, if $u^p_{n,t} = 1$, constraint (3.5c) states that a handover should happen from the serving BS p to a target BS candidate q in $\mathcal{Q}$. On the other hand, if $u^p_{n,t} = 0$, the constraint (3.5d) assures the UE status will remain unchanged. The latter relates to the case where a handover opportunity from BS p exists, but UE was not served by the BS p for τ or more timeslots earlier. In this case, there is no scope for such a handover. These two constraints tie the handover behavior because their left-hand sides are dual, e.g., when one of them is false the other constraint will force its right-hand side without making the model infeasible.

Moreover, constraint (3.5e) guarantees that the UE association is restricted to only one BS. A direct implication of this constraint is that there will be no more

than T timeslots, respecting the optimization observation period. This constraint's right-hand side might be changed to a constant that limits the maximum number of simultaneous connections instead of 1 to allow connection with multiple BS. By doing so, constraint (3.5b) must be changed to the constraint (3.7).

$$\sum_{m \in \mathcal{M}} \mathcal{R}_m \mathcal{S}^m_{n,t} x^m_{n,t} \geq C_n, \ \forall \ n \in \mathcal{N}, \forall \ t < T \tag{3.7}$$

3.2.2 Objective Function

The model is a framework that can be tailored for different objective functions depending on the design goals. We designed three objective functions to match the baseline and conditional handover protocols, and to optimize the most common metrics of QoS, capacity, delay and reliability.

Capacity Maximization

The baseline handover has only one subjacent utility function, to maximize the instantaneous RSRP and the SNR, by consequence. By maximizing the SNR, we also maximize the capacity, since by Shannon's capacity law SNR and capacity have a direct relation if the bandwidth is constant.

$$f(m, n, t) = \sum_{m,n,t} \mathcal{C}^m_{n,t} x^m_{n,t} \tag{3.8}$$

where $\mathcal{C}^m_{n,t}$ is the UE n Shannon's theoretical capacity with the BS m at the timeslot t and $x^m_{n,t}$ is the decision variable that defines if the UE n is associated to the BS m at timeslot t.

Joint Capacity Maximization and Delay Minimization

Maximizing SNR, or SNR-related metrics (SINR, capacity, data rate, etc) is usual in wireless network optimization [21, 83, 84]. We penalize the objective function, in order to minimize the message delay. Hence, the penalization increases with the message delay. As we aim to maximize the objective function, the best solution will avoid large penalization values to increase the final objective function value.

To model delay, we define a binary variable $y^m_{n,t}$. The variable $y^m_{n,t}$ is equal to 1 if the latest data message was sent at a timeslot t before the expiration of the UE delay tolerance. Delay tolerance, in this thesis, means the maximum latency the UE application can endure. Respectively, $x^m_{n,t}$ and $y^m_{n,t}$ represent the UE association to a BS and the low-latency requirements one needs to satisfy for applications such as autonomous vehicles.

Additionally, all UEs must have a delay tolerance requirement $\mathcal{D}_n$. The delay tolerance applies to the K_n data messages the UE and BS exchange. The k-th message arrives at time $A_k^n$and must be sent inside the delay window $\mathcal{W}_k$, as defined in Eq. (3.9).

$$\mathcal{W}_k = [A_k^n, A_k^n + \mathcal{D}_n]\,,\ \forall\ k \in K_n \tag{3.9}$$

$A_k^n + \mathcal{D}_n$ is the ultimate timeslot to send the k-th message. For the k-th message, the latency is equal to the difference $t - A_k^n$, which is equal to the current timeslot t, when the message is about to be sent, minus the message arrival timeslot A_k^n, similarly to the latency used in [85]. However, if $t > A_k^n + D_n$, then $y_{n,t}^m = 0$, because the message delay exceeded the maximum tolerable delay. Thus, the penalization for delaying the k-th message of user n at the timeslot t is given by Eq. (3.10).

$$\mathcal{I}_{n,t} = A_k^n + \mathcal{D}_n - t,\ \forall\ t \in \mathcal{W}_k. \tag{3.10}$$

The worst case for the objective function is when $\mathcal{I}_{n,t} = 0$, and the best case is when $\mathcal{I}_{n,t} = D_n$. By implication, $\mathcal{I}_{n,t} \in [0, D_n]$. Hence, the maximization will induce $\mathcal{I}_{n,t}$ to assume larger values, and consequently, the message transmission time t approximates the message arrival time A_k^n, minimizing the delay. The objective function in Eq. (3.11) jointly maximizes the capacity $\mathcal{C}_{n,t}^m$ while minimizing the delay. Due to the difference of magnitude between $\mathcal{C}_{n,t}^m$ and $\mathcal{I}_{n,t}$, this two constants must be normalized before the model solving.

$$f(m, n, t) = \sum_{m,n,t} \mathcal{C}_{n,t}^m x_{n,t}^m + \mathcal{I}_{n,t} y_{n,t}^m \tag{3.11}$$

To minimize the delay we need to add constraints to the problem in addition to the ones described in Section 3.2.1. First, the association and the delay variables need to be coupled. The message transmission is subject to the UE association to a BS and the capacity demand. Then, the constraint (3.12) accounts for the delay and association.

$$x_{n,t}^m \geq y_{n,t}^m,\ \forall\ n \in N,\ \forall\ m \in M,\ \forall\ t < T \tag{3.12}$$

Aiming to prioritize opportunities to exchange data during the delay window $\mathcal{W}_k$,

every message sent outside this window must not count for the objective function, i.e., $y_{n,t}^m = 0$. To keep the delay constraint valid only during the tolerance window, constraints (3.13) and (3.14) respectively impose that $y_{n,t}^m \leq 1$ within $\mathcal{W}_k$ and $y_{n,t}^m = 0$ otherwise.

$$\sum_{m \in M} \sum_{t \in \mathcal{W}_k} y_{n,t}^m \leq 1, \ \forall \ n \in N, \forall \ k \in K_n \tag{3.13}$$

$$\sum_{m \in M} \sum_{t \notin \mathcal{W}_k} y_{n,t}^m = 0, \ \forall \ n \in N, \forall \ k \in K_n \tag{3.14}$$

Joint Capacity and Reliability Maximization

As we stated in Section 2.2.3, the conditional handover minimizes Radio Link Failure (RLF) and handover failures by decoupling the preparation and execution phases. Hence, the earlier preparation eliminates dependency on the SBS and multiple BSs can be monitored in between the preparation and execution phases thereby improving the reliability. RLF avoidance depends on the BS channel conditions, i.e., BSs which can provide higher RSRP are unlikely to have the SINR below the out-of-synchronism threshold, Q_{out}. Thus, we define the variable $v_{n,t}^m$ that indicates whether the SINR $\mathcal{S}_{n,t}^m$ of the current BS is below Q_{out}. The variable $v_{n,t}^m$ is defined by Eq. (3.15).

$$v_{n,t}^m = \left(max\left\{sign\left(Q_{out} - \mathcal{S}_{n,t}^m\right), 0\right\}\right) x_{n,t}^m \tag{3.15}$$

In the outage condition, we have $\mathcal{S}_{n,t}^m < Q_{out}$, thus the sign and the max function will result in $+1$, then $v_{n,t}^m = x_{n,t}^m$. Otherwise, if $\mathcal{S}_{n,t}^m > Q_{out}$ the sign function will be equal to -1 but the max function will be equal to 0, then $v_{n,t}^m = 0$. Additionally, if the BS m is not selected as the source BS, then $x_{n,t}^m = 0$ and $v_{n,t}^m = 0$. We can write the objective function penalizing the timeslots in which the UE is in outage condition and $v_{n,t}^m = 1$. We can adjust the penalization by setting a constant $\kappa \in \{0, 1\}$. The objective function is described in Eq. (3.16).

$$f(m, n, t) = \sum_{m,n,t} \mathcal{C}_{n,t}^m \{x_{n,t}^m - (1 + \kappa) v_{n,t}^m\} \tag{3.16}$$

Due to the maximization sense, as in the joint capacity maximization and delay minimization, the problem solver will avoid high penalizations. Therefore, the system reliability is maximized by reducing the timeslots in outage conditions. This

differs from the capacity-only objective functions because the variable v intensifies the effect of the low channel quality timeslots which may lead to a solution that takes different handover decisions to avoid such timeslots.

3.2.3 MILP Model Final Remarks

Finally, the MILP formulation leads to the best handover decisions possible considering the different objective functions. We show in Appendix A that the computation time for solving this problem by enumeration grows with the simulation time T and the number of BSs $|\mathcal{M}|$ making it impossible to solve in feasible time for any instance of interest. Thus, this formulation is a good alternative to have an upper bound although it has drawbacks that might not permit it to work as a real handover mechanism. Due to the time needed to solve the problem and the necessity of having all the CSI beforehand, the MILP formulation presented can not work in real-time applications such as autonomous vehicles and Beyond Field of View UAVs. Thus we propose a heuristic to approximate the MILP solution by using limited time blockage predictions and dynamic programming.

3.3 Heuristic Handover

In this section, we will present the heuristic handover algorithm. Due to the blockage predictions, we assume that a different network architecture is needed in order to offer support for sensors, sensor data collection, and blockage prediction calculations. Then, the handover algorithm is described, from how the blockage predictions are processed into CSI to how the handover decisions are made and turned into a handover schedule.

3.3.1 System Architecture

Blockage prediction in wireless networks demands side information to feed the prediction algorithm. A myriad of side information can be used to leverage intelligence in wireless networks. This information can be related to the UE mobility, such as the position, velocity, acceleration and direction of the UE [20], originated from sensors widely available in devices nowadays. It is also easy to obtain channel information, like past in-band and lower bands CSI [74]. Multi-media information can also be used despite the higher cost and data size, such as images from cameras [86, 87], Light Detection and Ranging (LIDAR) scans [17, 30], Radio Detection and Ranging (RADAR) sensors [88]. These are examples of side information that the prediction algorithm can use.

Our goal is not to propose a blockage prediction algorithm. Rather, we are going to exploit the capabilities of blockage prediction algorithms already proposed and validated in the literature [17, 30]. We assume that the blockage prediction algorithm generates a limited length blockage prediction with length $\mathcal{L}$, a tuple of 0's and 1's flagging when there will be or not a blockage episode. Hence, we can say that a prediction done at timeslot t will contain the blockage prediction up to the $t + \mathcal{L}$ timeslot. It is easy to reason that the prediction's accuracy decreases with time, i.e., the prediction of t_i is more accurate than the prediction of t_j, since $i < j$.

Thus, we assume the predictions are generated based on scans from the LIDAR sensor equipped on the BSs and side information with the UE mobility data, such as velocity and position. These sensors periodically scan the BS surroundings and the BSs send the scans to a Central Processing Unit (CPU). Inspired by the Cell-Free Massive MIMO architecture [89] and based on the cloud support provided by the 5G [90] the CPU is a piece of equipment that can easily integrate the network infrastructure. The CPU's key role is to gather the LIDAR scan data and turn it into a prediction. This offloads the processing load from the BSs connected to the CPU via fronthaul link. The scans are comprised of the BS identification, the scan timestamp, the height of the scan and for each angle scanned the distance to the closest surface. After collecting the scans of all BSs, the CPU creates a temporal map that has the blockage status for each position covered and for the future $\mathcal{L}$ timeslots.

Not every BS might embed a LIDAR sensor and report its LIDAR scans to the CPU. Therefore, the BSs broadcast their LIDAR scan capability using 5G standard capability messages. Once the UE knows that the BSs nearby embed LIDAR sensors and support blockage prediction, the UE switches its operation mode to use the blockage prediction-based heuristic handover instead of baseline or conditional handover. In the blockage prediction support mode, the UE periodically requires a prediction to the SBS, which relays the requirement to the CPU. The requirement contains side information, such as UE identification, position, current speed, and direction of movement. With the UE position, speed and direction, the CPU can query the prediction for the current and future UE positions, generating a tuple of blockage prediction, as aforementioned. Only the BSs inside the range threshold R_{Th} are considered in the blockage prediction and prediction evaluation. This tuple is the input to an algorithm that evaluates the blockage predictions and schedules the handovers to come based on these predictions. There is a list of scheduled handovers for each UE, but only the handovers from the newest predictions persist on the list. Each scheduled handover is signalled to the SBS and TBS. When the time to perform a handover is about to come, the SBS sends a *RRC Handover Reconfiguration* to the UE, which disconnects from the SBS and connects to the TBS, as a

normal handover. In the next subsection, we will describe the handover protocol.

3.3.2 Heuristic Handover Protocol

Figure 3.1 depicts the heuristic handover operation. The normal function of BSs includes periodic LIDAR scans sent to the CPU. Nowadays, LIDAR sensors are at a few thousand dollar price-point and offer hundreds of meters range and high frame rate in outdoor environments [91]. The CPU gathers the scans and generates/updates the blockage predictions, which depend on the time ahead and the position. For each position covered, given a minimum resolution, there is a blockage prediction for each timeslot $\mathcal{L}$ that is updated every time new scans are available. The UE periodically requests a blockage prediction. The UE request period is longer than the CPU's prediction generation period. Thus, the CPU evaluates only the latest available blockage prediction using the dynamic programming algorithm.

Once the handover schedule is generated, the CPU negotiates the handover with the TBSs sending a *RRC Handover Request*. If there are resources available to receive the new UE, the TBS replies with a *RRC Handover Request Response*. This is similar to the baseline handover starting the handover preparation phase but with the CPU negotiating instead of the SBS. If accepted by the TBS, the CPU sends the handover schedule to the SBS, which will hold this until the time to execute comes or the handover is cancelled due to a newer prediction. When the execution time is due, the SBS sends the *RRC Handover Reconfiguration* to the UE, which will detach from the SBS and reconfigure itself to synchronize with the TBS. Henceforth, the handover protocol follows the same synchronization procedure as the baseline and conditional 5G handover, with the RACH procedure and the handover completion.

With the handovers scheduled, minimizing the effects of outage and HIT is possible. The handover decision-making algorithm may consider the relative position of the UE to the cell edge and anticipate an outage, avoiding handover failures and synchronization losses. The algorithm may schedule the handover in a way that the handover interruption period overlaps a predicted blockage episode to maximize the utility function. Knowing the TBS also helps decrease the effect of HIT, because it is possible to know when is the closest RACH or even make a contention-free RACH.

3.3.3 Dynamic Programming Handover Algorithm

Dynamic programming is a heuristic technique which accumulates the solution of sub-problems in a table, which is first checked every time a new sub-problem needs to be solved. By consulting the table, calculations may be skipped saving time and making the optimization problem more efficient. Dynamic programming has famous applications in Dijkstra's shortest path algorithm, integer knapsack problem, and

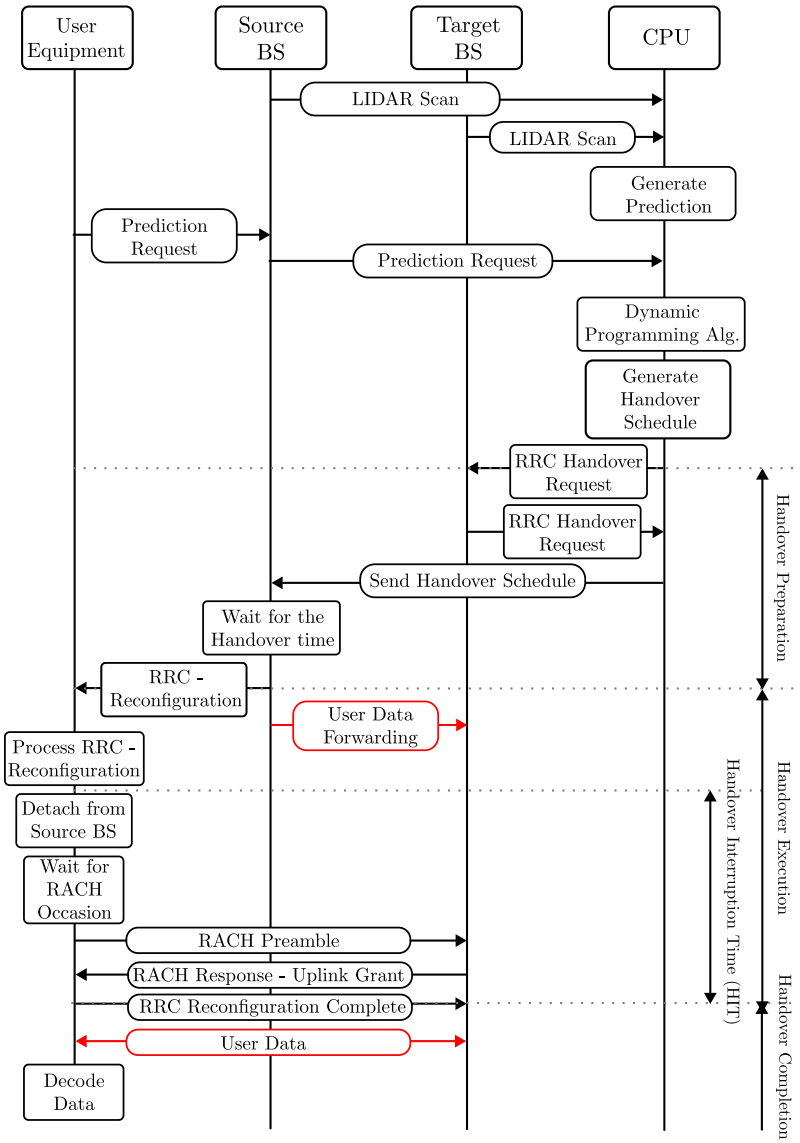


Figure 3.1: Heuristic handover protocol

tower of Hanoi puzzle solution. The CPU is responsible for executing the dynamic programming algorithm. After executing the algorithm, the handover decisions are output as a handover schedule with the BSs and timeslot at which each scheduled handover must be carried out.

The proposed handover algorithm receives as input the blockage prediction $\mathcal{B}_{\text{pred}}(m, n, t)$ for each BS m in the subset $\mathcal{M}' \subset \mathcal{M}$, which always includes the current SBS, for the UE n, given the UE position at timeslot t. The subset $\mathcal{M}'$ includes all the BSs no farther than R_{Th}plus the SBS. The predictions of length $\mathcal{L}$ are divided into smaller parts of length ℓ, then $\ell < \mathcal{L}$. At each ℓ interval the utility function $g(t, m, n)$ in Eq. 3.17 is evaluated.

$$g(m, n, t) = \sum_{k=t}^{t+\ell} \mathcal{C}\,(m, n, k)\, x_{n,k}^{m} \tag{3.17}$$

Where $\mathcal{C}(k, m, n)$ is the Shannon capacity calculated for the link between the BS $m \in \mathcal{M}'$ and the UE n using the predicted blockage status at k, $\mathcal{B}_{\text{pred}}(m, n, t)[k]$, as given by Eq. (3.18).

$$\mathcal{C}\,(m, n, k) = \mathcal{R}_m(SC_\eta \times 12) \log_2 \left(1 + \mathcal{P}_{n,k}^{m} - \sigma^2_{\text{noise},dB}\right) \tag{3.18}$$

Where $\mathcal{R}_m(SC_\eta \times 12)$ is the BSs bandwidth, as detailed in Subsection 3.2.1, $\mathcal{P}_{n,k}^{m}$ is the received power based on the path loss model in Eq. (3.2). The path loss model considers the blockage status and distance to the SBS, so the path loss attenuation for LoS is smaller than for NLoS for the same distance. Consequently, the value of $\mathcal{B}_{\text{pred}}(m, n, t)$ must be known to calculate the predicted capacity $\mathcal{C}(k, m, n)$. The utility function is valued for all the BS combinations in $\mathcal{M}'$, and a penalization p_{ho} is considered when changing from BS m to BS m'. The penalization represents the Handover Interruption Time (HIT) and forces the algorithm to place the handover in a moment when the HIT will impact the UE less. The goal of the penalization is not to inhibit handovers but to schedule the HIT without compromising the handover decision.

Therefore, we want to solve an optimization problem that is a simplified version of the optimization problem in Eq. (3.5). With less information, the problem in Eq. (3.19) does not consider the user packets' arrival and delay, and does not assure a capacity level. The objective function is the utility function in Eq. (3.17), which is the aggregated capacity evaluated in length $\mathcal{L}$ of a prediction and for all the BSs in the subset of BSs $\mathcal{M}'$, which includes the SBS plus the BSs in the range R_{Th}. The constraint (3.19b), defines the binary decision variable $w_{n,t}^{m}$, which is equal to 1 if a

handover will begin at the timeslot t and finish at the timeslot $t + p_{ho}$, and is equal to 0 otherwise. If $w^m_{n,t}$ is equal to 1, then one of the remaining BSs in $\mathcal{M}'$, except m from which the UE is leaving, will receive the UE by handover at $t + p_{ho}$, as in constraint (3.19c). Until the handover is finished, none of the BSs contribute to the utility function value and if there is no handover the BS status must remain the same, as described in constraints (3.19d) and (3.19e), respectively. The constraints (3.19f) and (3.19g), assure that no more than one BS will be counted at the same time for the solution and that the optimization always starts by the UE's SBS.

$$\mathcal{P}2 : max \quad g(m, n, t) \tag{3.19a}$$

$$\text{s.t.} \quad w^m_{n,t} = \max\left\{(x^m_{n,t-1} + \sum_{m' \in \mathcal{M}' \setminus \{m\}} x^{m'}_{n,t+p_{ho}}) - 1, 0\right\}, \tag{3.19b}$$

$$w^m_{n,t} = 1 \rightarrow \sum_{m' \in \mathcal{M}' \setminus \{m\}} x^{m'}_{n,t+p_{ho}} \leq 1, \tag{3.19c}$$

$$w^m_{n,t} = 1 \rightarrow \sum_{m \in \mathcal{M}'} \sum_{k=t}^{t+p_{ho}} x^m_{n,k} = 0, \tag{3.19d}$$

$$w^m_{n,t} = 0 \rightarrow x^m_{n,t-1} = x^m_{n,t}, \quad t \leq \mathcal{L}, \quad \forall m \in \mathcal{M}', \tag{3.19e}$$

$$\sum_{m \in \mathcal{M}'} x^m_{n,t} \leq 1, \quad \forall m \in \mathcal{M}', \tag{3.19f}$$

$$x^m_{n,0} = 1, \quad m = BS_{\text{source}}, \tag{3.19g}$$

$$(3.19b), (3.19c), (3.19d), \quad t \leq \mathcal{L} - p_{ho}, m \in \mathcal{M}' \tag{3.19h}$$

The algorithm starts at the SBS and assesses whether staying at the SBS or handing over to another BS in $\mathcal{M}'$ will maximize the utility function value. If there is no SBS, e.g., after a Radio Link Failure (RLF) has happened or the UE has just entered a blockage prediction support region, all the BSs in the subset $\mathcal{M}'$ are considered starting BSs. The Algorithm 1 details the procedure to find the best handover scheduler using blockage predictions. From lines 3 to 7, the algorithm evaluates the predicted Shannon capacity, as in Eq. (3.18), and stores the values into a tuple. Next, the algorithm tests if there is a defined SBS or if it will need to define a new SBS. If an SBS already exists, the algorithm from lines 11 to 12 will find the maximum value for the utility function starting the search from the SBS. Otherwise, in lines 14 to 20, the algorithm will test all the BSs in the subset $\mathcal{M}'$ to find the starting BS that leads to the highest utility function value. Finally, with the global utility function value found, in line 22 the algorithm generates the handover schedule that resulted in the highest utility function value.

In lines 12 and 15 of the Algorithm 1, the function in the Algorithm 2 is called.

Algorithm 1: Heuristic handover algorithm based on predictions

Function `PredictionHandover`($\mathcal{B}_{pred}(m,n,t)$, $\mathcal{M}'$, $\mathcal{L}$, ℓ, t, BS_{source})
  $\mathcal{C} \leftarrow \{\}$
  for *each BS* m *in* $\mathcal{M}'$ **do**
    for *each* k *in* $\{t,\ \mathcal{L}\}$ **do**
      $\mathcal{C}\,(k,m,n) \leftarrow W \log_2\left(1+\mathcal{P}^m_{n,k}-\sigma^2_{\text{noise},dB}\right)$
    end
  end
  memory $\leftarrow \{\}$
  global_utility_value $\leftarrow\ -\infty$
  if BS_{source} *exists* **then**
    best_starting_bs $\leftarrow\ BS_{\text{source}}$
    global_utility_value $\leftarrow$ `DynamicProgrammingHandover`($\mathcal{C}$, $\mathcal{M}'$, $\mathcal{L}$, ℓ, t, BS_{source}, *memory*)
  else
    for *BS* m *in* $\mathcal{M}'$ **do**
      local_utility_value $\leftarrow$ `DynamicProgrammingHandover`($\mathcal{C}$, $\mathcal{M}'$, $\mathcal{L}$, ℓ, t, m, *memory*)
      if *local_utility_value* $>$ *global_utility_value* **then**
        global_utility_value $\leftarrow$ local_utility_value
        best_starting_bs $\leftarrow\ m$
      end
    end
  end
  handover_schedule $\leftarrow$ `HandoverScheduler`(*memory,best_starting_bs*)
  return *handover_schedule*
end

This is a recursive function to find the highest utility function value. The recursion ends if $t \geq \mathcal{L}$, as shown in lines 2 to 4, because there is no more predicted capacity to be calculated. Also, if the value of this recursion is already in the recursion table, called *memory*, the function returns this value and avoids doing the calculation another time, as presented in lines 5 to 7. Then, the algorithm starts the search for the maximum utility function value. The algorithm will test all the combinations of BSs for every ℓ timeslots and whenever the BS changes, the algorithm skips p_{ho} timeslots. If the BS is still the same, the function is called again for the same BS m for the timeslots from $t_0 + \ell$, as detailed in line 12. Otherwise, as shown in line 14, if the BS changes from m to $next\_BS$, then the function is called again for the BS $next\_BS$ for the timeslots starting from $t_0 + p_{ho}$. Then a recursion call is initiated for each BS in $\mathcal{M}'$, starting from t_1 equals either $t_0 + p_{ho}$ or $t_0 + \ell$, whether it was called to evaluate a different BS or not, respectively. The $|\mathcal{M}'|$ recursions will then evaluate the next $t_1 + \ell$ timeslots for the BS $next\_BS$ passed as argument, and each one will call more $|\mathcal{M}'|$. The k-th recursion call of the BS $m \in \mathcal{M}'$ will finish the recursive loop if $t_k > \mathcal{L}$, returning 0 to the utility value of the t_{k-1} recursion call, which will return $g(t_{k-1}, m, n)$ for the t_{k-2} recursion, until recursion t_0 is finished. In this case, the Algorithm 2 was called a total of $|\mathcal{M}'|^k+1$ times.Finally, in lines 16 to 17, the achieved local maximum utility function value is compared to the current maximum utility function value, and in line 19 the maximum value is stored in the memory tuple.

The other function called from Algorithm 1 is the function that generates the handover schedule, described in Algorithm 3. This function is similar to the one in Algorithm 2, but its goal is to find the handover schedule that leads to the maximum utility function value found by Algorithm 2. Thus, in lines 8 to 12, this function traces back the utility function values found by Algorithm 2, which are now in the *memory*. If the value found is equal to the maximum utility function value, the BS is added to the handover schedule, as presented in lines 13 to 18. Every step of the algorithm re-evaluates each handover opportunity, but the computational cost is minimal as the utility function's values were already calculated and stored in the memory when Algorithm 2 ran. Then the BSs and the handover timeslots which led to the maximum utility function value are included in the solution.

Considering that the UE will stay connected to the SBS for the whole prediction length, the utility function is evaluated $\frac{\mathcal{L}}{\ell}$. Assume that, in the worst case, the UE stays associated with the current BS for ℓ timeslots. Thus, if there will be one handover to any of the $|\mathcal{M}'|-1$ target BSs after these first ℓ timeslots and then there will be no other handovers, the utility function is evaluated $\ell + \frac{\mathcal{L}-(p_{ho}+\ell)}{\ell}$. We can ignore the first ℓ timeslots because they are already stored in the dynamic programming table, reducing the previous formula to $\frac{\mathcal{L}-(p_{ho}+\ell)}{\ell}$. For each extra

Algorithm 2: Handover dynamic programming algorithm

```
1  Function DynamicProgrammingHandover(C, M', L, ℓ, t, m, memory)
2      if t ≥ L then
3          return 0
4      end
5      if (t,m) in memory then
6          return memory(t, m)
7      end
8      max_utility_value ← 0
9      utility_value ← −∞
10     for each next_BS in M' do
11         if next_BS = m then
12             utility_value ← g(t, m, n) + DynamicProgrammingHandover(C,
                M', L, ℓ, t + ℓ, m, memory)
13         else
14             utility_value ← g(t, m, n) + DynamicProgrammingHandover(C,
                M', L, ℓ, t + p_ho, next_BS, memory)
15         end
16         if utility_value > max_utility_value then
17             max_utility_value ← utility_value
18     end
19     memory(t, m) ← max_utility_value
20     return max_utility_value
21 end
```

Algorithm 3: Handover scheduling algorithm

```
1  Function HandoverScheduler(C, M', L, ℓ, t, m, memory)
2      if t ≥ L then
3          return {}
4      end
5      max_utility_value ← memory(t, m)
6      utility_value ← −∞
7      for each BS next_BS in M' do
8          if next_BS = m then
9              utility_value ← g(t, m, n) +DynamicProgrammingHandover(C,
                M', L, ℓ, t + ℓ, m, memory)
10         else
11             utility_value ← g(t, m, n) + DynamicProgrammingHandover(C,
                M', L, ℓ, t + p_ho, next_BS, memory)
12         end
13         if utility_value = max_utility_value then
14             if next_BS = m then
15                 return {(m, t)} + HandoverScheduler(C, M', L, ℓ, t + ℓ,
                    m, memory)
16             else
17                 return {(m, t)} + HandoverScheduler(C, M', L, ℓ, t + p_ho,
                    m, memory)
18             end
19     end
20 end
```

Table 3.2: Time complexity mean and standard deviation for different prediction windows

Prediction window length	Dynamic Programming		Optimization	
	Mean	Standard Deviation	Mean	Standard Deviation
500	**0.009604**	0.000168	0.805352	0.051471
1000	**0.015472**	0.000008	1.092928	0.216898
1500	**0.023195**	0.000100	1.201409	0.403058
2000	**0.031159**	0.000167	1.211698	0.584794

handover, the interval to evaluate the utility function reduces by $(p_{ho} + \ell)$. Hence the maximum number of handovers is equal to $\frac{\mathcal{L}-\ell}{p_{ho}+\ell}$, totalling $\mathcal{G}$ realizations of the utility function $g(m, n, t)$ in Eq. (3.17), as given by Eq. (3.20)

$$\mathcal{G} = \frac{\mathcal{L}}{\ell} + (|\mathcal{M}'|-1) \sum_{i=1}^{\frac{\mathcal{L}-\ell}{p_{ho}+\ell}} \frac{\mathcal{L} - i(p_{ho} + \ell)}{\ell} \tag{3.20}$$

Each realization of $\mathcal{G}$ is the sum of ℓ values stored in the tuple $\mathcal{C}(m, n, k)$. Considering the complexity of accessing the position in a tuple of a known element equal to 1, the dynamic programming algorithm's worst-case complexity is equal to $\ell\mathcal{G}$, given by Eq. (3.21).

$$\mathcal{O}\left\{\mathcal{L} + (|\mathcal{M}'|-1) \sum_{i=1}^{\frac{\mathcal{L}-\ell}{p_{ho}+\ell}} \mathcal{L} - i(p_{ho} + \ell)\right\} \tag{3.21}$$

The complexity in Eq. (3.21) only depends on the heuristic handover parameters, namely the number of target BSs $|\mathcal{M}'|$, the prediction length $\mathcal{L}$, the dynamic programming step ℓ, and the handover penalization p_{ho}. Thus, these parameters can be tweaked to fit these necessities in case of restricted computational resources and computational time shortage. The dynamic programming complexity grows linearly with $\mathcal{M}'$ and $\mathcal{L}$ compared to the brute force complexity in Eq. (A.4) that presents polynomial growth if we used $\mathcal{M}'$ and $\mathcal{L}$ instead of $|M|$ and T, respectively. Besides, on average, dynamic programming executes faster than MILP optimization problems for instances of the same size. Table 3.2 displays the time complexity mean and standard deviation comparing the dynamic programming and the optimization for different prediction windows in timeslots, for velocity equal to 30 km/h and blockage density equal to 0.003. We notice that the dynamic programming computes the handover decision in less time than the optimization for the same prediction window length.

This handover technique has some minor pitfalls we must account for before comparing it to other handover techniques. The first thing that may be of concern is the accuracy of predictions. Because of the limited time knowledge, even with perfect predictions, the dynamic programming algorithm is less efficient than the optimization in its handover decision. So, we might expect a decrease in the heuristic handover performance when the blockage prediction algorithm accuracy decreases. Additionally, sending the scans and periodically requesting predictions may congest the fronthaul, and although the prediction request is short, its frequency should not overburden the fronthaul. The scans' frequency and resolution must likewise be set carefully so as not to bottleneck the fronthaul links. Alleviating this might be possible by adapting the prediction request period and the LIDAR frame rate and resolution to the surrounding situation. Analogue to video Codification Decodification (CODEC) techniques, if the surrounding scenario is not changing much, then the LIDAR frame rate and resolution may decrease, for instance.

Another point is the architecture and protocol modifications necessary for the blockage prediction-based heuristic handover to work. We consider a CPU connected to all BSs in a region, so if the number of BSs increases we might have a scalability problem. Although it is not the goal of this thesis, few works on the CF-MIMO area approach this problem, like [92] but still, there is ground to cover in this area of research. The CPU itself is more and more envisioned as part of the network as the Open Radio Access Network (OpenRAN) architecture evolves and virtualization becomes integrated into the telecommunication field [90]. We also present a modification in the handover protocol but they may be supported by already existing handover protocol messages, as depicted in Fig. 3.1. If the CPU acts as a central agent with controlling capabilities that could in real-time manage these Layer 3 messages, such as in Near Real-Time RAN Intelligent Controller (Near-RT-RIC), then the heuristic handover may be incorporated into a working 5G network [93, 94].

Nonetheless, the architecture to support this handover is feasible and resembles promising enabling technologies, such as OpenRAN and CF-MIMO. Besides, the heuristic is based on the main features of the optimization, which are using future information and optimising a KPI under the handover restrictions, but with lower computational complexity and feasible in real-time. The prediction and handover decision algorithms play key roles in the heuristic handover. The prediction algorithm must provide accuracy and low complexity whereas the handover decision algorithm should approximate the optimization decision avoiding unnecessary handovers. If these conditions are true, the heuristic handover technique will be efficient and enable strict QoS applications.

3.4 Conclusion

In this chapter, we presented our two main contributions: the MILP formulation for handover and KPI optimization and the heuristic handover algorithm based on blockage predictions. The MILP optimization is formulated to integrate handover conditions and capacity and delay level constraints. The MILP objective function can be changed to optimize different KPIs. We propose three objective functions optimizing capacity, capacity plus delay assurance, and capacity plus reliability optimization. Derived from the optimization and exploiting the concept of knowing CSI from future timeslots we also propose a heuristic algorithm to make the handover decisions. The algorithm uses blockage predictions to estimate a predicted capacity and schedule handovers to maximize a utility function.

Reproducing the handover conditions as an optimization problem constraint is not trivial. The handover decision must be made during the problem-solving and depends on the association decision variables x. Thus, the solver must only include a handover at t in the solution if, for the same BS p, $x^p_{n,k} = 1$ for $k = t - \tau - 1, t - \tau, \ldots, t - 1$, where τ is the time to trigger, and another BS q RSRP is better than p by a margin in the same period. As multiple conditions are involved, it is not trivial to turn this into part of the optimization problem and to the best of our knowledge our model is a novel contribution. To overcome this, we first enumerated all the possible handover opportunities, i.e., every timeslot a BS RSRP is greater than another BS RSRP by a margin, the corresponding β is equal to 1. From β we create the set of BS candidates $\mathcal{Q}$ and the variable u constraints in Eq. (3.4). Note that, $\tau - \sum_{j=t-\tau-1}^{t-1} x^p_{n,j}$ is equal to 0 only if the BS p is the SBS of UE n for more than τ timeslots, and is the key condition to put a handover on the solution. Then, the conditions to handover considering the objective function are given by the constraints (3.5c) and (3.5d). The preprocessing and the indicator constraints improved the model efficiency decreasing the time needed to solve instances.

Though it offers the optimal solution, the drawbacks of the optimization model are the need for all CSI information at the beginning of the optimization and the time it takes to compute the solution for an instance. The heuristic handover circumvents these two problems by solving a simplified version of the optimization problem, in smaller chunks of time in a restricted set of BSs, using short-term future blockage predictions, that are feasible using the LIDAR sensor's side information. Dynamic programming is used to iterate over the BSs' blockage prediction and find the best moments to make a handover and schedule it. Different from a brute force or exhaustive search algorithm, dynamic programming exploits the utility function values already stored on the table. However, this approach required the conception of a protocol and architecture to communicate the blockage predictions and

handover decisions between all the entities efficiently. We consider these changes a minor problem in our proposal because 5G NR handover does not operate this way by default. However, the modifications we propose to the handover are either supported in OpenRAN or are well-accepted in the literature as future technological improvements.

In the next chapter, we introduce the simulation scenario and the results of the handover methods. The assessment of the handover techniques is divided into two parts, the evaluation of the handover parameters impact and the effects of velocity and blockage density. This helps us find points of improvement and strengths that can serve as guidelines for designing new handover techniques and evaluating them.

Chapter 4

Numerical Evaluation and Discussion

In this chapter, we detail the system model and present the results of comparisons between the 3GPP 5G NR baseline and conditional strategies with the optimization model and heuristic handover technique, detailed in Chapter 3. In Section 4.2, we introduce the simulation environment for each one of the handover techniques assessed in this chapter. In Section 4.4.1, we will discuss the performance of the baseline and conditional handover methods regarding the handover parameters. Finally, we make our final remarks by analyzing the performance gaps between the algorithms and the upper bound set by the optimization model.

4.1 System Model

The system model adopted in this thesis consists of a set M of mmWave BS serving a set N of UE. Henceforward, we use m to refer to a single BS and n to refer to a single UE. Each BS has $\mathcal{R}_m$ Physical Resource Blocks (PRB) with subcarrier bandwidth φ to offer. Then, for every Transmission Time Interval (TTI) of duration δ, one UE is granted with the frequency domain resources available at the BS it is connected to. The values of δ and φ for 3GPP 5G NR are determined by the current numerology in use [95].

We adopt the shadowing path loss model from [82], which increases the path loss exponent and the reference distance path loss in the case of NLoS. The carrier frequency picked is 28 GHz, because it offers lower atmospheric absorption, rain attenuation [34], and better overall Signal-to-interference-plus-Noise Ratio (SINR) [96] than higher mmWave carrier frequencies, such as 73 GHz. However, any mmWave carrier frequency would work with our optimization model. The path loss equation for LoS and NLoS cases is estimated as in Equation (3.2) [82].

In addition to the path loss model, a multipath fading model is adopted to reflect the effect of mobility. As the UE moves, the signal interacts with the scenario obstacles and is decomposed into multipath components that interfere constructively

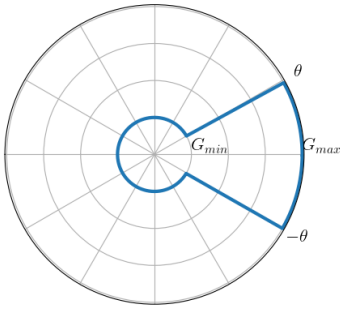


Figure 4.1: Sectored antenna model radiation pattern

and destructively with each other and the LoS component. Specifically, we use the Clarke's "sum-of-sinusoids" algorithm to generate Rayleigh fading components variation over time for narrowband channels. The RSRP measurements use the DMRS signal contained in the SS Blocks, spread over 4 OFDM symbols, and for the effect of handover decision and RSRP measurements sampling, we can consider a flat-fading method due to the narrow channel.

Besides, we consider a radiation pattern that consists of a main and a secondary lobe. This model is called the sectored antenna model [97]. The radiation pattern of the sectored antenna model is illustrated in Figure 4.1. The antenna main lobe beam width is θ_b, and G_{max} and G_{min} are, respectively, the antenna gain in the main and the secondary lobes, as expressed in Eq. (4.1).

$$G(\theta) = \begin{cases} G_{max}, & \text{if } |\theta| \leq \theta_b \\ G_{min}, & \text{otherwise} \end{cases} \tag{4.1}$$

In the system model, some form of interference is assumed. Using the same path loss model in Eq. (3.2), all the neighbouring BSs generate interference over the source BS and UE link. The interference defines the UE synchronization status, packets and control messages transmission success, and whether a handover will fail or be completed. Next, we will introduce the simulation scenario.

4.2 Simulation Details

In this section, we will describe the implementation of the simulation scenario. All simulations had the instances as a common starting point. These instances contain a description of the scenario and network devices. The scenario contains information such as the noise power, frequency of operation, and the random number generator seed. The network device classes are divided into base stations and user equipment.

The base stations are characterised by their subcarrier spacing, the number of available resource blocks, and location in the scenario. The user equipment description comprises initial coordinates in the scenario, velocity in x and y axis, number of sent packets, and each packet arrival timeslot. Finally, the instances also store the blockage status of each pair UE-BS, for each timeslot and the location of the blocking obstacles.

From these input instances, generated for different seeds, velocities and scenario blockage densities, each algorithm pre-processes the instances, computes its handover decisions and post-processes these decisions to output the KPIs. The preprocessing functions of each simulation only differ in the output, the system model applied is the same. The simulations differ in how the handover decisions are computed. The baseline and conditional handovers were implemented in a discrete event simulator, thus the handover decisions are made as the events scheduled happen. The optimization requires that all CSI are available before the optimization starts and the model is written according to the variables and pre-processed CSI. The heuristic uses the same pre-processing of the optimization but iterates over the CSI step-by-step. In the following section, we will describe these simulations in more detail and the challenges of implementing each one.

4.2.1 3GPP 5G NR Baseline and Conditional Handover

For the 5G baseline and conditional methods, a Python discrete event simulator with more than 9300 lines of code implements the handover and PHY Layer functions. The code was written using interface definitions that can be modified or extended to support different implementations of objects, for example, mobility models, user equipment functionalities, handover mechanisms, and CSI measurement and analysis. The discrete event simulator is configured to work with timeslots of 1 ms. In this case, the preprocessing loads the instance and creates the mobile UE and the BS objects. When the simulation starts, some UE and BS services are initiated. The UE has the mobility model service, that updates the UE position, the measurement services, and the data packet transmission service. The BS initiates with the Synchronization Signal (SS) and Random Access Shared Channel (RACH) services.

The UE does the Layer 1 and Layer 3 measurements with 20 ms and 100 ms period. For every new Layer 3 measurement available, the UE assesses the necessity of making a handover and if the UE stays in synchronicity with its Source BS. These periodic services repeat during the simulation and some information is logged, such as the Reference Signal Received Power (RSRP), SNR and SINR levels, the BSs the UE was associated with and when the UE began and finished each association. At the end of the simulation, with the UE association and the measurements the UE

collected and stored during the simulation, the KPIs are calculated and output.

The synchronicity evaluation follows the procedure described in Section 2.2.2, with the out-of-synchronicity threshold $Q_{out} = -8\ dB$, the in-synchronicity threshold $Q_{in} = -6\ dB$, the number of in-synchronicity samples N310 = 1, the number of in-synchronicity samples N311 = 1, and the synchronicity evaluation timer T310 equal to 600 ms. If the UE is found out-of-synchronicity, it is disconnected from the SBS and starts the connection reestablishment process detailed in Section 2.2.2.

The handover mechanism used by the UE has a specified interface, thus it is possible to plug in the baseline handover class or conditional handover class to the same UE object, preserving the remaining simulation code. The BSs that meet the A3 event handover conditions generate a handover process that follows the lifecycle defined in the 3GPP standard for either the baseline or the conditional method. The conditional differs from the baseline in the handover execution phase, when the conditional handover monitors all the BSs in the execution list analysing the execution and release condition, while the baseline carries on the handover with the BS with the best RSRP when the Measurement Report is sent to the SBS. Handover failures occur in the simulation if a handover control message cannot be sent due to low SINR, if the UE loses connectivity with the SBS or if the RACH procedure with the TBS cannot be completed before the T304 = 500 ms timer expires.

4.2.2 Mixed Integer Linear Programming and Heuristic Handover

The Mixed Integer Linear Programming (MILP) formulation was implemented using the Gurobi solver [98] Python library. To make the code more efficient we pre-process the RSRP, SNR and SINR, and list all the possible handover opportunities that satisfy the A3 handover event conditions. Thus, the pre-processing code iterates at every 1 ms, the same as the baseline and conditional simulation. In each iteration of the pre-processing, the RSRP, SNR and SINR are calculated assuming the same channel model and Layer 1 and Layer 3 filters as the baseline and conditional simulations. This way, the only difference between the optimization and the baseline and condition simulation, is that the optimization accesses all the CSI beforehand when the model is built.

The pre-processed data are stored in tuples that serve as input to the code routine that creates the MILP model. Another routine creates the variables, the constraints, and the objective functions, as described in Eq. (3.5), and adds to the model. At the end of the optimization solving the post-processing iterates over the model variables' final values and obtains the UE association to the BSs. With the association and the CSI, such as SINR, the KPIs can be calculated and output.

Likewise, the heuristic handover does the same pre-processing of the optimization, but without enumerating the handover opportunities. The CSI output at the end of the pre-processing is only used later in the post-processing to evaluate the KPIs. In the main code, the UE position is updated and the handover scheduler is checked every 1 ms, and for each 500 ms a prediction request is sent. When a prediction request is sent, the blockage predictions are collected and evaluated using the dynamic programming algorithm. We assume that the blockage prediction algorithm has inaccuracies and that the accuracy decreases with the prediction length increase.

In our evaluation, we assume a burst error function over the blockage predictions. To decide whether there will be an error we use an exponential error function, described by Eq. (4.2). The value of $p(t)$ is the parameter of a Bernoulli distribution to determine if the current predicted timeslot will be predicted correctly or not. The burst length is given by a Poisson distribution with parameter $\lambda_B = 10$. After one of these error functions is applied to the BSs in range predictions, the dynamic programming algorithm finds the best handover schedule and updates the handover scheduler using the utility function in Eq. (3.19). The scheduler only considers the handovers scheduled from the newest blockage predictions, due to the decreasing accuracy of the latter predicted timeslots. At the end of the simulation, the post-processing is made from the UE association with BSs equal to the optimization.

$$\varepsilon(t) = a + be^{\frac{t*b}{100}},\ a = 0.014,\ b = 0.086 \tag{4.2}$$

There are difficulties in implementing the heuristic handover in the same framework as the baseline and conditional handover methods. Some of these challenges could not be solved within the thesis's timeframe. Because of its proactive nature that triggers the previously scheduled handovers, a new UE implementation is the easiest way to not conflict with the other handovers implementations, that react to a Layer 3 measurement event. Another problem to be solved is to remove older handovers from the scheduler which requires cancelling all other routines linked with these handovers. These challenges were circumvented with a dedicated discrete event simulation code different from the one described in Subsection 4.2.1 without forfeiting it as well. However, we envision in future work implementing the heuristic handover algorithm in the same simulator where the baseline handover and the conditional handover algorithms were implemented. Thus, we also suggest a new BS implementation with a data structure supporting tracking each scheduled handover. Table 4.1 describes the parameters used in the heuristic handover simulations.

Table 4.1: Heuristic Handover Parameters

Parameter	Symbol	Value
Prediction Length	$\mathcal{L}$	1000
Algorithm step	ℓ	10
Handover interruption	p_{ho}	100
Range of candidate BSs	R_{Th}	300 m

4.2.3 Simulation Scenario

In our simulations, we aim to reflect a vehicle-to-infrastructure scenario, where the vehicle travels and communicates with the BS via a mmWave radio interface, which is possibly blocked by static obstacles. Three handover mechanisms will be compared in this chapter: the 3GPP baseline handover, the 3GPP conditional handover and the heuristic handover, alongside the optimization model with the joint capacity and reliability maximization objective function. The latter with the joint capacity and reliability maximization objective function is comparable to the conditional handover, as both optimize reliability, i.e., minimize outage. We found in our tests and assessments that $\kappa = 0.4$ provided the best results for handover optimization with the joint capacity and reliability maximization objective function.

Hence, we conduct Monte Carlo simulations over 60 sample instances for each parameter's value combination and all the plots show 95% confidence intervals around the mean. We evaluate the handover mechanism implementations by varying the UE velocity and the blockage density (λ). There are $|M|$ BS in the scenario, deployed on a hexagonal grid with BS radius ρ. The obstacle positioning follows a Homogeneous Poisson Point Process (HPPP) with parameter (λ) and changes at every scenario realization resulting in different LoS conditions and, consequently, different BS RSRP.

Regarding the mobility model, we consider the UE as a vehicle moving on a straight route with path length L and constant speed v. We call this the constant speed model. Though we adopted such a model, it does not represent a limitation for either the proposed optimization model or the 5G NR baseline handover simulation. Actually, the optimization relies only on the UE-BS SNR values at each timeslot, so any mobility model or trace can be plugged into the optimization model. As mentioned in section 2.2.2, the TTT and the UE velocity are inversely proportional. Thus, as the velocity increases, the time to react to the faster channel changes decreases, so the TTT must be reduced. In this paper, we consider the following velocity-TTT pairs: (30 km/h, 160 ms), (60 km/h, 80 ms), and (90 km/h, 40 ms). These values are based on [5], to reproduce a low and high-velocity scenario and the correspondent TTT. Different TTT values for the same values of velocity, if too high

or too low, might result in signal quality worsening or excessive handovers [99, 100]. Additionally, we consider a Poisson distributed packet arrival with mean λ_p and packet length of γ bits, including overhead and payload, which refers to a 24 Frames Per Second (FPS) 720p video frame. We summarise the simulation parameters in Table 4.2.

Table 4.2: Simulation Parameters [4, 5]

Parameter	Symbol	Value
TTT	τ	$[160, 80, 40]$ ms
Velocity	v	$[30, 60, 90]$ km/h
Blockage Density	λ	$[0.001, 0.003, 0.005, 0.007, 0.009]$ objects/m^2
Delay Tolerance	$\mathcal{D}_n$	$[2, 4, 6, 8, 10]$ ms
TTI	δ	1 ms
SCS	φ	60 kHz
Packet Rate	η	$\frac{1}{120}$ packets/ms
Packet Length	γ	2025 kB
Noise Power	σ_n	-100 dBm
BS Radius	ρ	150 m
Number of BS	$\lvert M \rvert$	13 BS
Path Length	L	$1.2\sqrt{2}$ km
L1 Sampling Period	ω_{L1}	20 ms
L3 Sampling Period	ω_{L3}	100 ms
L3 filter Coefficient	k	1
Out-of-sync Threshold	Q_{out}	-8 dB
In-of-sync Threshold	Q_{in}	-6 dB
N310 counter		1
N311 counter		1
T310 timer		600 ms
T304 time		500 ms

4.3 Evaluation Metrics

Evaluating the performance of handover mechanisms in wireless networks is necessary to ensure the appropriate network operation and seamless UE experience. Particularly when the QoS requirements are strict and the system is dynamic and presents randomness, a thorough evaluation of handover-related metrics, such as

ping-pong handovers and handover rate, and general KPIs, like SNR and capacity is necessary. This section outlines the metrics used to evaluate the handover performance in the proposed scenario of 5G NR mmWave network mobility under blockage.

- **Handover Success Ratio:** This metric measures how robust the handover mechanism is to channel variations. As stated in section 2.2.2, if the handover is too late or too soon, the handover will fail. Therefore, the higher the handover success ratio, the better the handover decision. The handover success ratio is the quotient between the number of handover successes and the total number of handovers, which is equal to the number of successes plus the failures, as in Eq. (4.3).

$$HO_{\text{success}} = \frac{N_{\text{success}}}{N_{\text{success}} + N_{\text{failure}}} \tag{4.3}$$

- **Handover Failure Ratio:** On the other hand, the handover failure ratio exposes the weakness of a handover mechanism or its parameters configuration. If a handover is too long, for instance, it is more likely to fail due to the distance to the SBS which increases the path loss. More details on the handover failure conditions are in Section 2.2.2. The handover failure ratio complements the handover success ratio, i.e., is equal to the number of handover failures divided by the total number of handovers, as described in Eq. (4.4).

$$HO_{\text{failure}} = \frac{N_{\text{failure}}}{N_{\text{success}} + N_{\text{failure}}} \tag{4.4}$$

- **Handover Rate:** The handover rate is a metric to evaluate the handover decision lifespan. A high handover rate reflects the handover mechanisms' poor decisions that are remedied with other subsequent handovers, reducing the Time of Stay (ToS) and increasing overhead and outage due to HIT. Thus, the handover rate is given by the total number of handovers divided by the simulation time, as expressed in Eq. (4.5).

$$HO_{\text{rate}} = \frac{N_{\text{success}} + N_{\text{failure}}}{T} \tag{4.5}$$

- **Time of Stay:** The time of stay (ToS) is the interval a UE stays associated with the same SBS. If the time of stay is long, this indicates that the handover decision to that BS was good. Otherwise, a short time of stay indicates that a handover could be skipped. This metric also defines whether a handover is considered ping-pong and together these two metrics are crucial to determine

if a handover mechanism provides adequate handover decisions. However, the time of stay also captures the time of staying when the connection between the UE and SBS is broken.

- **Ping-pong Ratio:** The ping-pong handovers are handovers considered unnecessary because the time of stay in BS is too short. A handover is considered ping-pong when the UE bounces between two BSs A and B and the ToS of the intermediary handover to BS B is shorter than the MToS. The ping-pong ratio is equal to the fraction of the number of well-succeeded handovers classified as ping-pong, as shown in Eq. (4.6).

$$HO_{\text{ping-pong}} = \frac{N_{\text{ping-pong}}}{N_{\text{success}}} \tag{4.6}$$

- **Out-of-sync Rate:** In section 2.2.2 we defined the conditions for an UE to be considered out-of-synchronicity with the SBS. It happens when a certain number of SINR samples are below Q_{out} without any sample above Q_{in} for a given period. The out-of-sync rate measures how many events like this happen in a second, indicating if the handover decisions are leading to good channel conditions so the UE can keep connected to the SBS.

- **Conditional Handover Preparation Ratio:** One important metric when assessing a handover method is how much overhead such a handover method generates. The conditional handover preparation phase requires the SBS to communicate with each TBS that meets the preparation condition before adding them to the execution list. Thus a high preparation ratio means that too many BSs are being prepared for each executed handover, generating excessive overhead.

- **Conditional Handover Release Ratio:** Similarly to the conditional handover preparation ratio, the release ratio measures how well-adjusted are the preparation and execution conditions. If many TBSs are released, either the preparation offset margin is too low such that BSs in bad channel conditions are being prepared or the execution offset margin is too high such that the BSs are being released because could not achieve such threshold.

- **SINR:** The SINR is defined as the ratio of the received signal strength to the interference and background noise level, as given by Eq. (4.7). The SINR defines the UE conditions regarding the synchronization, determining whether the UE is in RLF condition. Consequently, a UE with a high average SINR is unlikely to get in RLF conditions, handover failures and in need of connection

re-establishment.

$$SINR_{n,t}^{m} = \frac{10\log_{10}\mathcal{P}_{n,t}^{m}}{\sum_{i\in\mathcal{M}\setminus\{m\}} 10\log_{10}\mathcal{P}_{n,t}^{i} + \sigma^2_{\text{noise,linear}}} \tag{4.7}$$

- **SNR:** The SNR is the ratio of the received signal strength to the background noise level, detailed in Eq. (3.1). High SNR values indicate that the handover decisions favoured the UE to have high-quality data transmission and minimal error rates.

- **Capacity:** The capacity sets an upper bound on the data rate of a communication channel based on the bandwidth available at the physical resource blocks $\mathcal{R}_m$ and the SNR. Regarding the handover, the capacity allows us to assess the handover decisions, for example, whether the handovers lead to high attenuation and high interference BSs. With the insights the capacity provides, we can evaluate the efficiency and reliability of handover processes, and its ability to deliver seamless connectivity for mobile users.

- **Outage:** The outage is the sum of all the intervals where the UE did not have a data connection with the source BS. Hence, if the UE SINR is below Q_{out} we consider it an outage interval. The HIT and connection re-establishment intervals are also considered outage intervals.

- **Delivery Ratio:** We consider that a packet can only be sent if the UE is currently associated with an SBS and the reception is good if the SINR is above Q_{out}. Otherwise, consecutive retransmission attempts follow until the channel conditions improve. If a new data message arrives and the previous message has not been transmitted yet, the older one is discarded. Let $\bar{K}_n$ be the number of discarded packets, then the delivery ratio is the fraction of the K_n packets of UE n that were successfully received, as given by Eq. (4.8).

$$R_{\text{delivery}} = \frac{K_n - \bar{K}_n}{K_n} \tag{4.8}$$

- **Delay-Constrained Delivery Ratio:** Similarly to the delivery ratio, the delay-constrained delivery ratio is the fraction of packets sent before the expiration of the delay tolerance window $\mathcal{W}_k$, as shown in Eq. (3.9). This metric is bounded by the delivery ratio because, in the best scenario, all packets successfully received arrive inside the window $\mathcal{W}_k$.

4.4 Results

We assess in this section the 3GPP 5G NR baseline and conditional handover, our proposed optimal handover policy, and the heuristic handover technique performances. We present the results in two different sets. First, we analyze the impact of the handover parameters, TTT and margin, on the baseline and conditional handover observing the gap to the MILP optimization model. Second, we compare the baseline, conditional, and heuristic handover methods with the optimization model to observe the speed and blockage density effects.

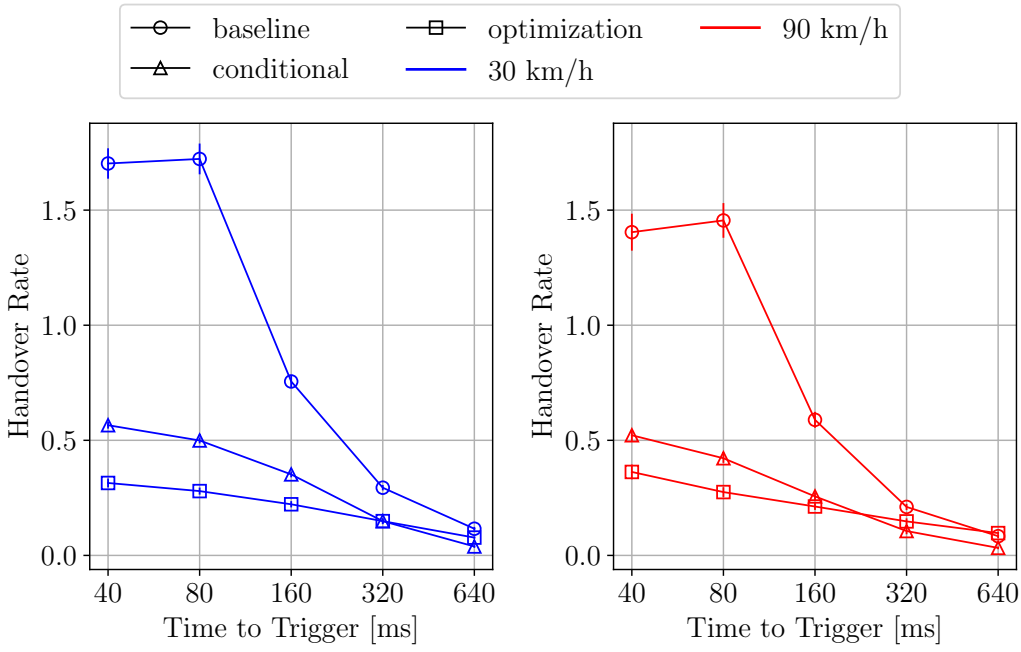
In each plot, we refer to the 3GPP 5G NR baseline and conditional handover techniques as baseline and conditional. The MILP optimal formulation for handover with the joint capacity and reliability maximization objective function, which is the upper bound, is referred to in the plots as only optimization. These are evaluated in both two sets of results, the one surveying the handover parameters in Subsection 4.4.1 and the other analyzing velocity and blockage density in Subsection 4.4.2.

Regarding the heuristic handover, the plots in Subsection 4.4.2 present the dynamic programming algorithm's results assuming blockage prediction error and perfect blockage prediction. The former is called heuristic with burst error, which uses the aforementioned function error described in Subsection 4.2.2, while the latter is called heuristic with perfect CSI, recalling that the algorithm turns the blockage prediction into predicted CSI. However, the curves of these two variations of the heuristic handover simulation overlapped in almost all plots. Indeed, the error function could not significantly affect the heuristic handover performance in the current parameter set-up and simulation scenario, which will be discussed later in Subsection 4.4.2.

4.4.1 Impact of Handover Parameters

In this subsection, we analyze the impact of the handover parameters, specifically TTT and offset margin, on the handover KPI. As highlighted in Section 2.2.2, there are works in the literature dedicated to optimizing such parameters. However, the goal of this subsection is not to optimize them but to demonstrate how these parameters influence the handover. We evaluate two main parameters, the handover RSRP margin, which is equal to the offset plus the hysteresis on the right-hand side of Eq. (2.4), and the TTT.

First, we analyze the impact of the TTT. The TTT changes the handover mechanism sensibility to the channel variations, which should adapt to the UE speed. In high-speed situations a shorter TTT allows the UE to react to power decay at the cell border when distancing from the SBS and approaching the TBS. Otherwise, the UE can cross the cell border with a delayed handover and handover control messages



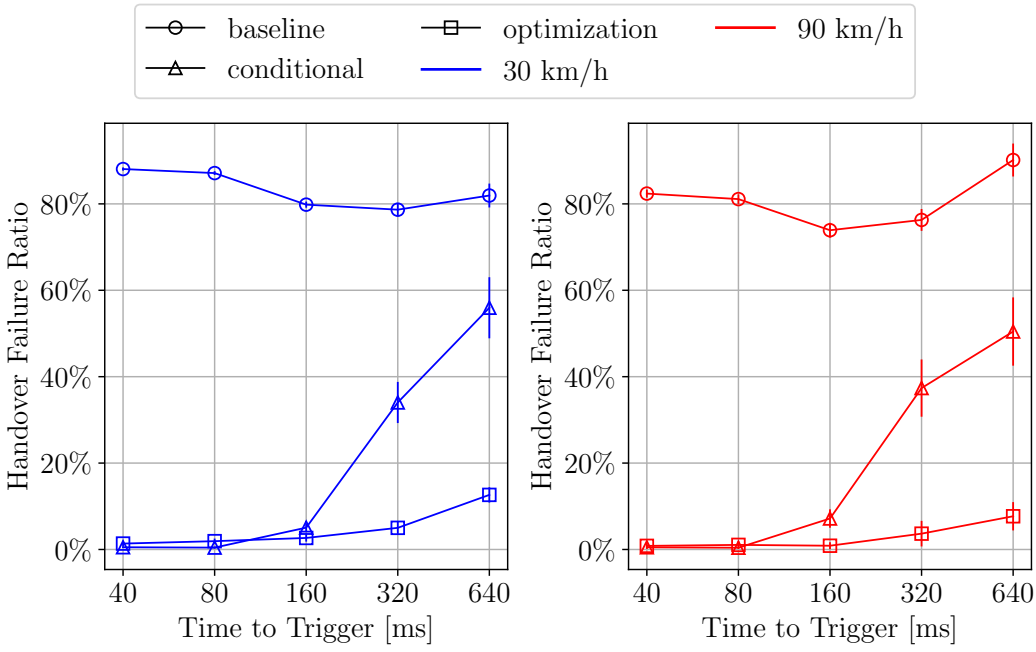
(a) 30 km/h (b) 90 km/h

Figure 4.2: Handover rate in handovers per second vs TTT

pending to be exchanged with the SBS, increasing the handover failure probability. On the other hand, if the TTT is too short for a low-speed UE, unnecessary handovers will be triggered, increasing the network load and the overhead. For this analysis, we fixed the blockage density at 0.003 obstacles/m^2 and the offset margin at 3 dB.

We evaluated the impact of the TTT on the handover algorithms. Primarily, all the mechanisms exhibit a decrease in the handover rate because of the longer TTT which makes the handover mechanism insensitive to the channel changes, as depicted in Fig. 4.2. If the TTT is too long, not even the growing distance to the SBS will trigger the handover and the UE will reach the cell edge without having a TBS to handover. We can project that in ideal conditions the handover rate would be 0.0733 and 0.2133 handovers per second for the velocities of 30 km/h and 90 km/h, respectively. This projection is the ratio between the cell diameter and the UE velocity. When the UE reaches the cell border it might handover to that BS and when it leaves at the diametrical point, considering a straight route, it might handover to a neighbour BS. Thus, depending on the UE velocity, the time to traverse a BS coverage area varies and, consequently, the handover rate. This ratio is underestimated because it does not consider the blockage episodes and the channel effects, such as large and small-scale fading, which are conditions that might impose a handover when the UE is still in the middle of the SBS coverage.

As a negative outcome of such a limited handover rate, the handover failure



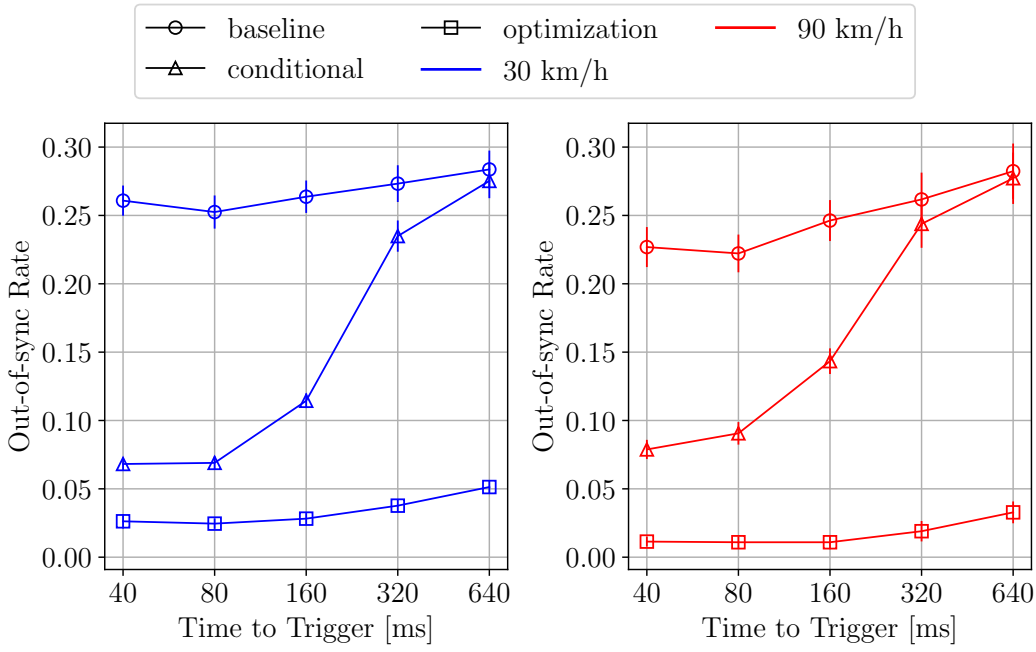
(a) 30 km/h (b) 90 km/h

Figure 4.3: Handover failure rate

increases with the TTT, as shown in Fig. 4.3. The longer TTT delays the handover decision making and the SBSs are likely too far from the current UE position which increases the probability that a handover control message will be lost. A flattening can be noticed for TTT values smaller than the Layer 3 measurement period in both the handover rate and the failure ratio. In this condition, it is unlikely that the RSRP measurement will be updated between the triggering condition and the measurement report transmission. Thus, when the time to make the handover decision comes, the RSRP sample is the same that triggered the handover. This evens out the handover events for TTT values shorter than the measurement period because the condition to make the handover decision does not change during the TTT.

Due to the low handover rate, the UE stays associated with the SBS for longer intervals. As the staying lasts longer the UE gets distant from the SBS causing the worsening of the channel quality. Figure 4.4 depicts the rate of out-of-synchronization episodes by the TTT. As the TTT increases the number of out-of-sync episodes also increases, indicating that the TTT is no longer adequate to maintain the connection with a BS that offers a good quality channel. The longer TTTs are causing a delay in the handover choice because the TBSs must keep the handover condition for a prolonged period and such condition may be interrupted by a blockage episode. Thus, the UE gets far from the SBS without having a TBS causing the loss of synchronization.

The conditional handover presents the steepest out-of-sync increase. Each of the



(a) 30 km/h

(b) 90 km/h

Figure 4.4: Number of out-of-sync episodes vs the TTT

two conditions, preparation and execution, requires a TTT. Therefore, it is unlikely that a TBS will hold both conditions for a longer TTT, especially the execution condition that usually has a higher RSRP offset above the SBS' RSRP. The delay in the execution condition is a known weakness of the conditional handover, which can be seen in Fig. 4.4 aggravated by the increasing TTT.

Increasing the TTT delays the handover decision and the conditional handover mechanism is more exposed to RLF with the SBS if the execution is delayed. Another negative effect is the growth of the conditional handover preparations in relation to the successful handovers, as detailed in Fig. 4.5. Such a rise of 57% and 59% for 30 km/h and 90 km/h, respectively, generates a high load in the network due to the preparations that were not executed and only increased the network overhead.

The margin of RSRP, comprised of the offset plus the hysteresis margins, must be set so the channel fluctuations do not unnecessarily trigger handovers. Regarding the conditional handover, which has more than one margin to be configured, when not mentioned we refer to the execution margin. The margin should not be too large inhibiting the handovers or too tight allowing slight channel changes to initiate a handover. If the margin is increased, the handover mechanism becomes more selective, allowing fewer handovers to happen, as the TBS' RSRP has to be greater than the SBS' RSRP by a larger value. The handover rate decreases as the margin increases due to the scarcity of BSs with channel conditions good enough to accomplish the handover event condition, as depicted in Fig. 4.6. The baseline exhibits significantly more handovers in the smaller margin configurations because

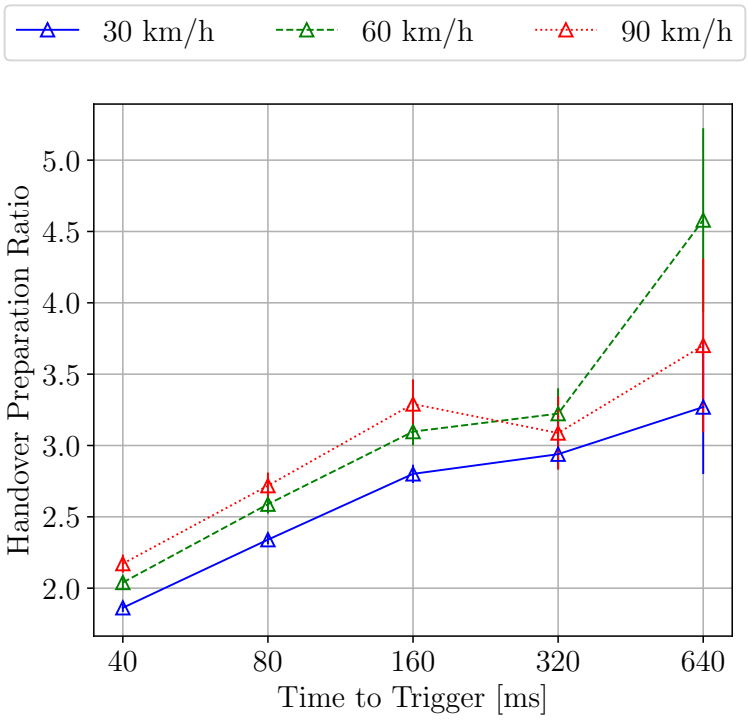
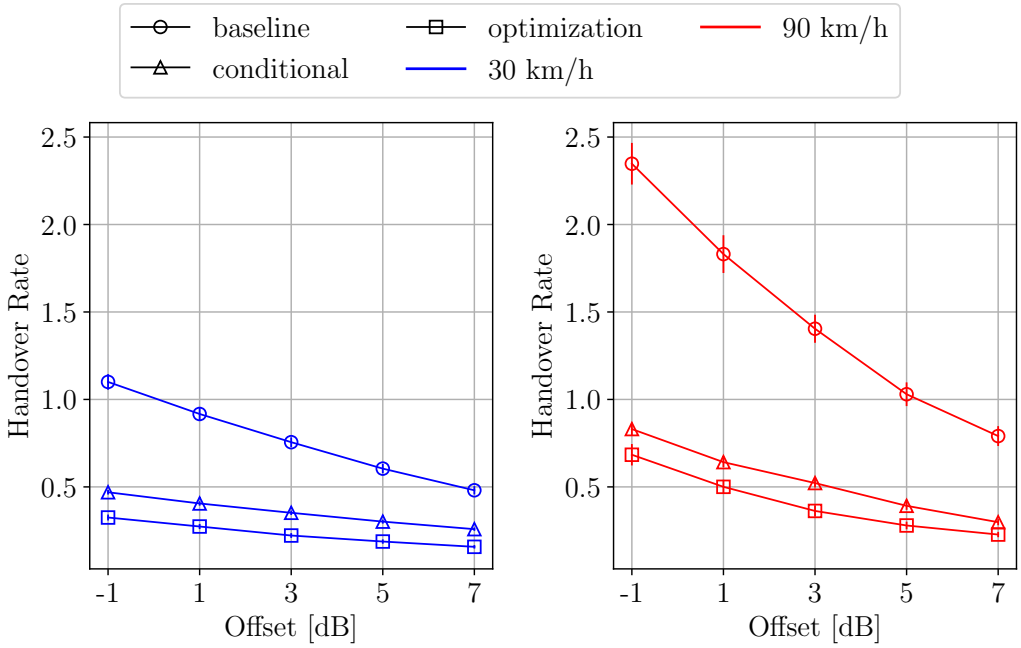


Figure 4.5: Conditional handover preparations by the time to trigger.

it is greedy in its decision and assesses the TBS with only one condition. Rather, the conditional handover has two conditions preventing greediness once the TBS must meet both conditions with individual TTTs.

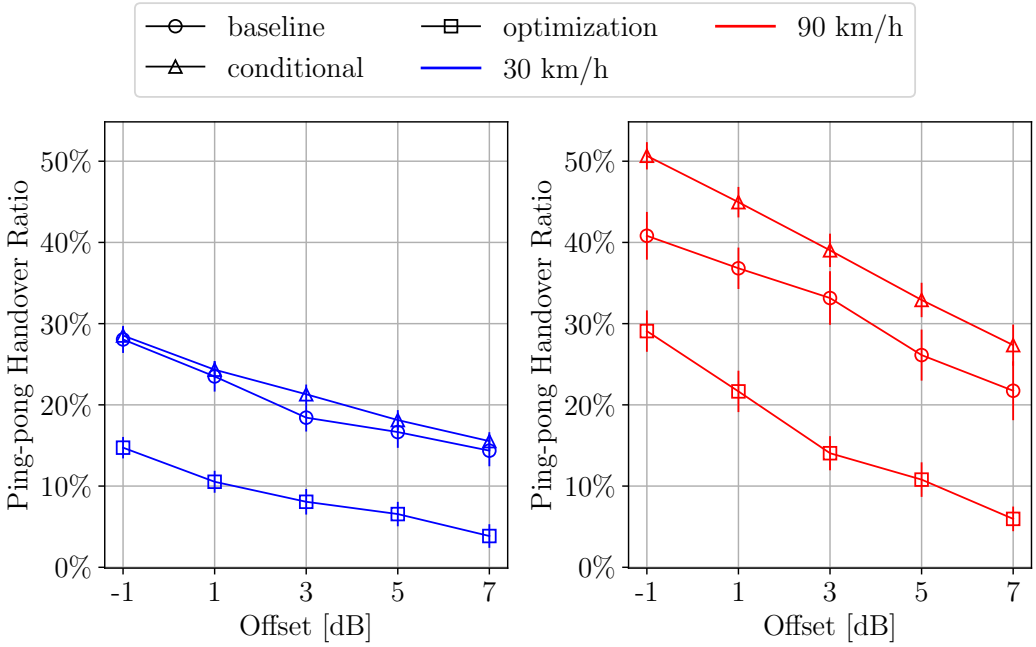
If the margin is short, the number of handovers increases once channel fluctuations may cause unnecessary handovers. Another evidence of a small margin is the number of ping-pong handovers. Ping-pong handovers are undesired because the handover to the intermediary BS is unnecessary and does not last the minimum time of stay. Besides, the overhead generated by the ping-pong handovers aggravates the network congestion and increases the time the user is under outage due to HIT. Figure 4.7 highlights that the ratio of ping-pong handovers decreases for larger RSRP margins, reducing the overhead and the HIT, that interrupts data connection, benefiting the user.

We may conclude from Fig. 4.7 that increasing the margin has only positive effects as the ping-pong handover ratio decreases, ideally what is desired from a handover mechanism. Due to the handover failures the baseline ratio of ping-pong handovers is lower than the ping-pong, once those ping-pong handover attempts can not be completed and counted as realized ping-pong handovers. Although increasing the margin has decreased the ping-pong handover, augmenting the margin reflects on the channel quality. As the handovers decrease the handover failures slightly increase, as depicted in Fig. 4.8. As the BSs able to surplus larger margin values are scarce, the handover opportunities reduce. Hence, the UE stays more time



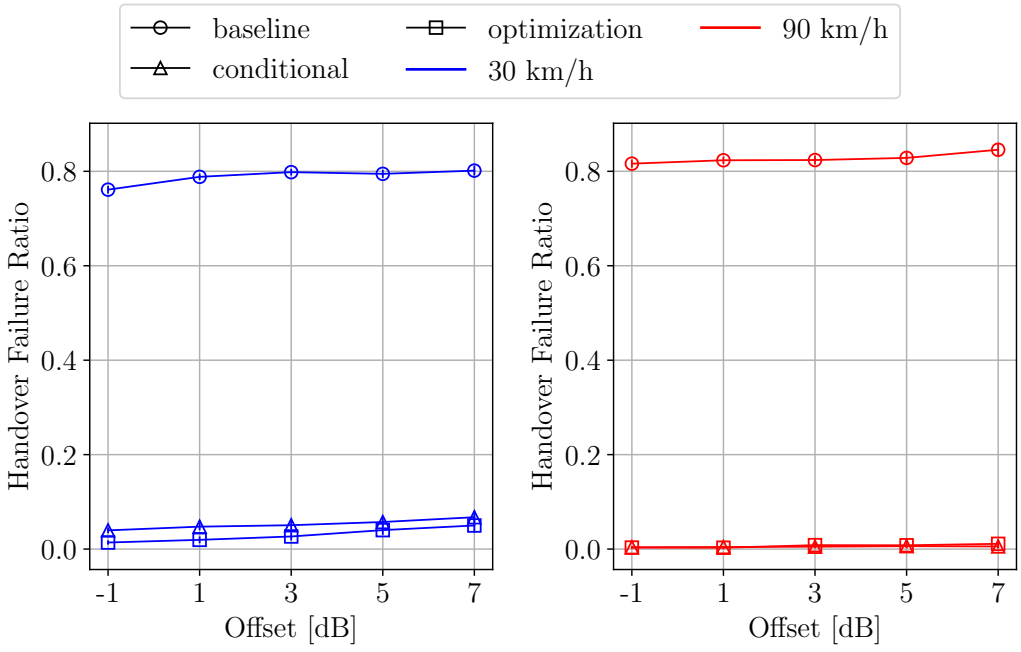
(a) 30 km/h (b) 90 km/h

Figure 4.6: Handover rate vs. the offset margins.



(a) 30 km/h (b) 90 km/h

Figure 4.7: Handover ping-pong ratio vs. the offset margins.



(a) 30 km/h (b) 90 km/h

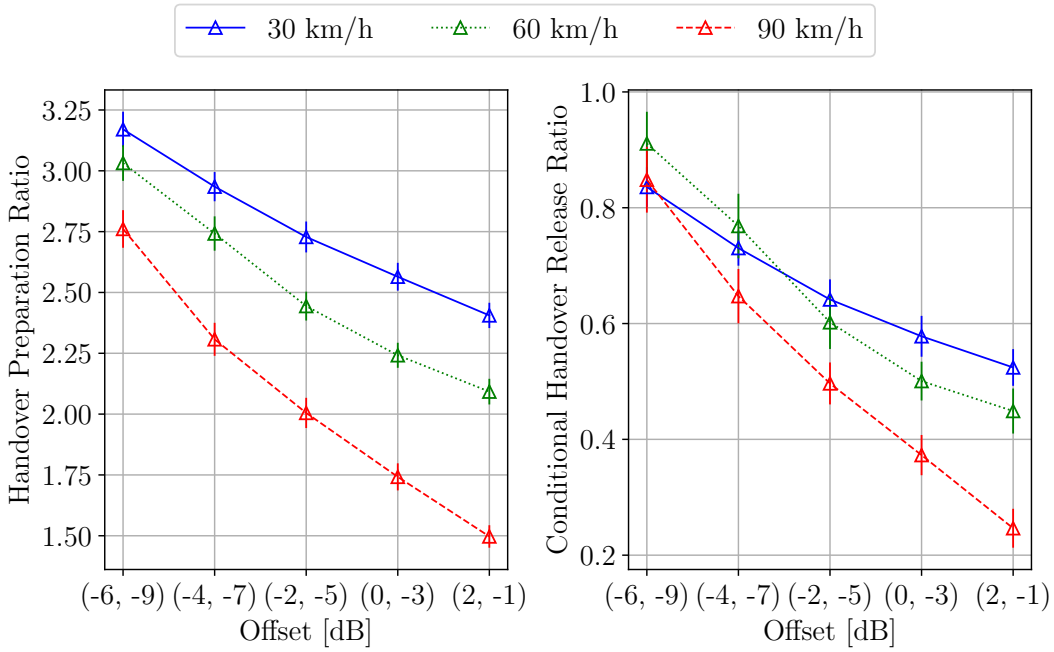
Figure 4.8: Handover Failure ratio vs. the offset margins.

associated with the SBS and experiences channel degradation with the distancing.

The conditional handover technique has two other offsets: the preparation and the release. The preparation offset sets the condition for a BS to be included in the execution list while the release offset determines the margin to stop the monitoring of a BS during execution and exclude the BS from the execution list. For the next results we fixed the execution offset at 3 dB and varied the preparation and release offsets in pairs ranging from -6 to -2 dB and -9 to -5 dB, for the preparation and release, respectively.

The preparation offset is analogous to the execution offset, the shorter the margin the more BSs will be prepared. Preparing many BSs increases the handover overhead due to the negotiation between the SBS and the candidate TBS before the TBS enters the execution list. Moreover, as the release offset must be equal to or less than the preparation offset, the probability of releasing the prepared TBSs grows as the release offset decreases for the same value of execution offset. On account of the lower preparation margin for the same execution margin, it is unlikely that the prepared TBSs will reach the higher execution offset and they can be rather released from the preparation list.

Fig. 4.9 displays the ratio of prepared and released BSs to successful conditional handovers. As previously stated, the lower the preparation/release the more the number of prepared/released BSs for a fixed execution offset margin. More than 80% of the prepared TBSs are released from the execution list for the lowest combination



(a) Conditional handover preparation ratio (b) Conditional handover release ratio

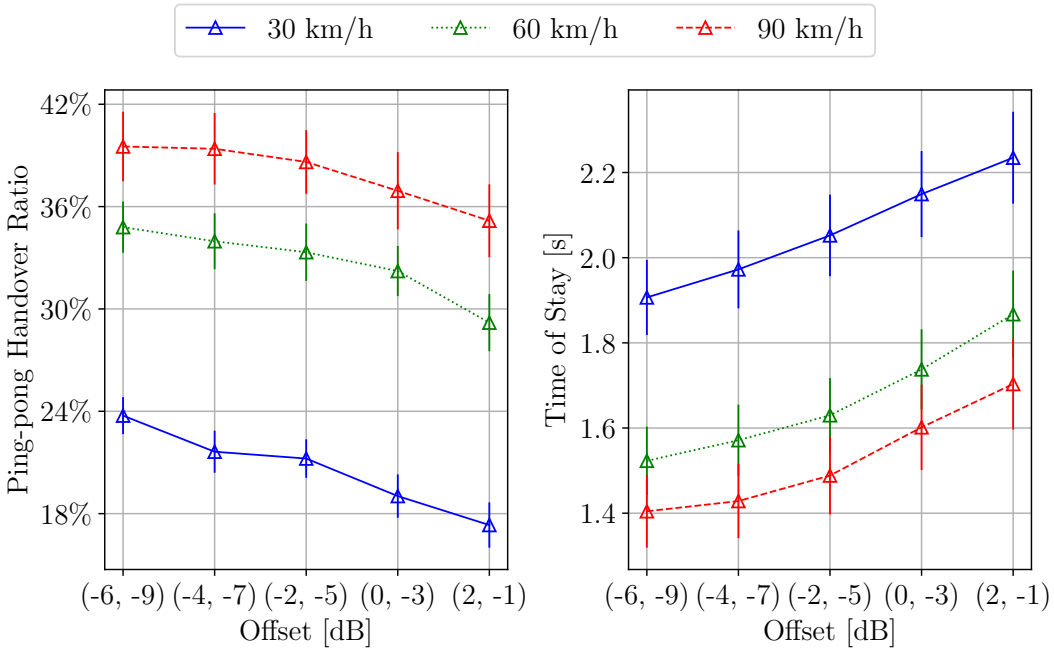
Figure 4.9: Conditional handover preparation ratio and release ratio of stay vs. the conditional handover preparation and release offset margins.

of preparation and release offset. This confirms that lower margin values for the conditional handover preparation and release offset are detrimental to the network operation. For lower speed, this effect is even worse, as it presents a less inclined slope than the high speed. When caught in a bad channel condition, such as a blockage, the UE takes more time to move and change this condition at a lower speed.

Finally, the ping-pong handover ratio and the time of stay have the same trend as the execution offset, though more subtle as Fig. 4.10 shows. We attribute this to a dominance of the execution offset margin over the preparation and release margins, which indeed determines whether the conditional handover will take place. Nevertheless, by choosing the highest values for the preparation and release margin the ping-pong ratio decreases and the time of stay increases. However, increasing those values also decreases the number of prepared BSs, reducing the handover options and jeopardizing the conditional handover operation.

4.4.2 Impact of Velocity and Blockage Density

The velocity and blockage are the defining factors for the channel condition the UE experiences. The velocity defines how fast the UE approaches and moves away from a BS. The blockage density determines how frequently a blockage episode happens or the likelihood of blockage episodes. Both parameters are also coupled to the



(a) Ping-pong handover ratio

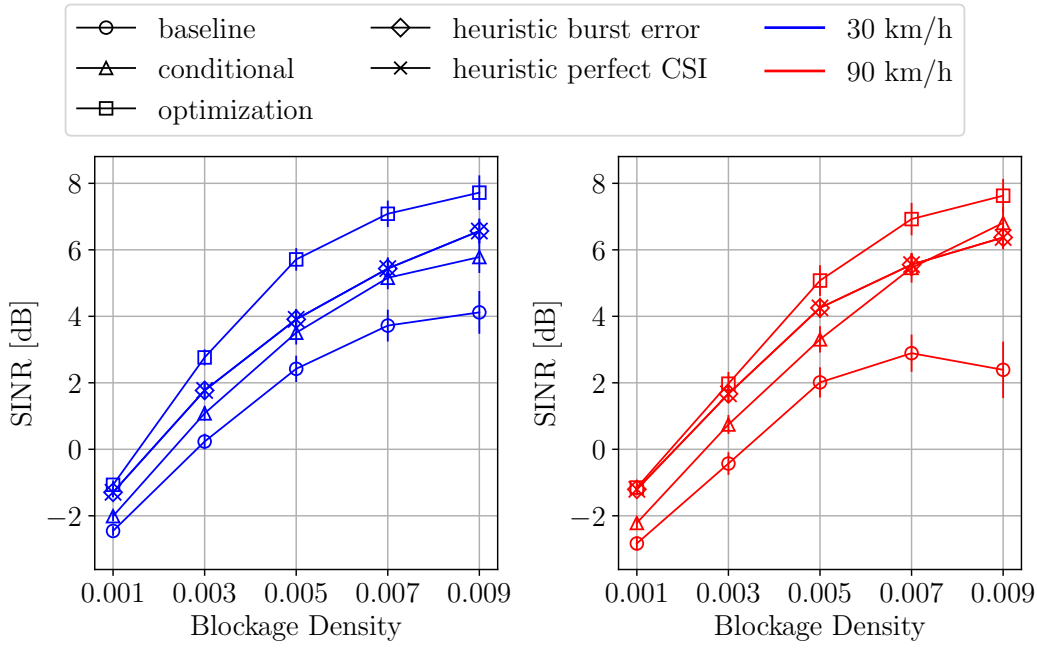
(b) Time of stay

Figure 4.10: Ping-pong handover ratio and time of stay vs. the conditional handover preparation and release offset margins.

blockage duration, the velocity is inversely proportional and the blockage density is directly proportional to the blockage duration. An example of such influence is the interference aggravation in lower blockage density situations because the interfering BSs are blocked and the sum of the interfering signal power is lowered by blocking obstacles.

In Fig. 4.11, the SINR growth is due to the greater probability that the interfering neighbour BSs will be blocked reducing the overall interference and augmenting the SINR. On the contrary, for lower blockage densities, it is unlikely that the neighbour BSs will be blocked, then the sum of the interference generated by these BSs increases and the SINR diminishes. Comparing the scenarios with low and high-velocity it is easy to notice a reduction in the difference between the algorithms. When the UE moves at a higher velocity, the blockage episodes tend to last less as the UE passes quickly by the objects. Therefore, it is more likely that the UE is at LoS situations when the velocity is high. Consequently, the interference generated by neighbour BSs is higher and the SINR is lower when the user is moving with high velocity. The baseline handover shows the worst performance among the algorithms, due to its greediness in the most challenging scenario of high blockage density and high velocity where its handover decisions do not lead the UE to good channel conditions.

On the other hand, the SNR follows the opposite trend, illustrated by Fig. 4.12, decreasing as the blockage density grows. As the SNR only depends on the received



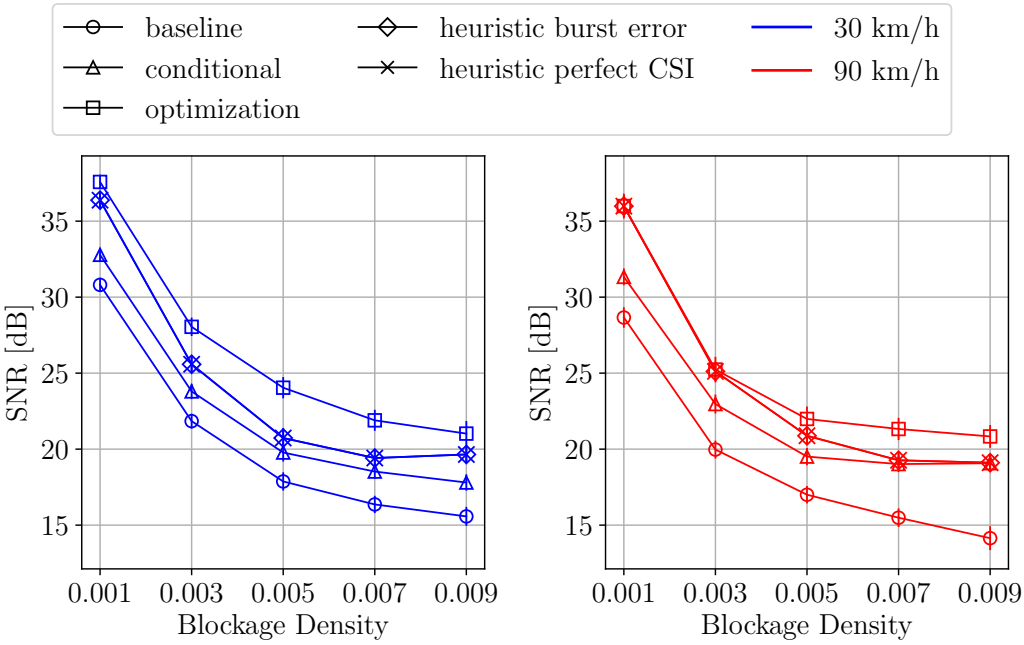
(a) 30 km/h (b) 90 km/h

Figure 4.11: SINR vs. blockage density

power and the noise power level, for the same noise power increasing blockage density decreases the overall received power samples. Since the SINR and SNR only differ by the interference in the quotient, this confirms how the blockage density shapes the interference.

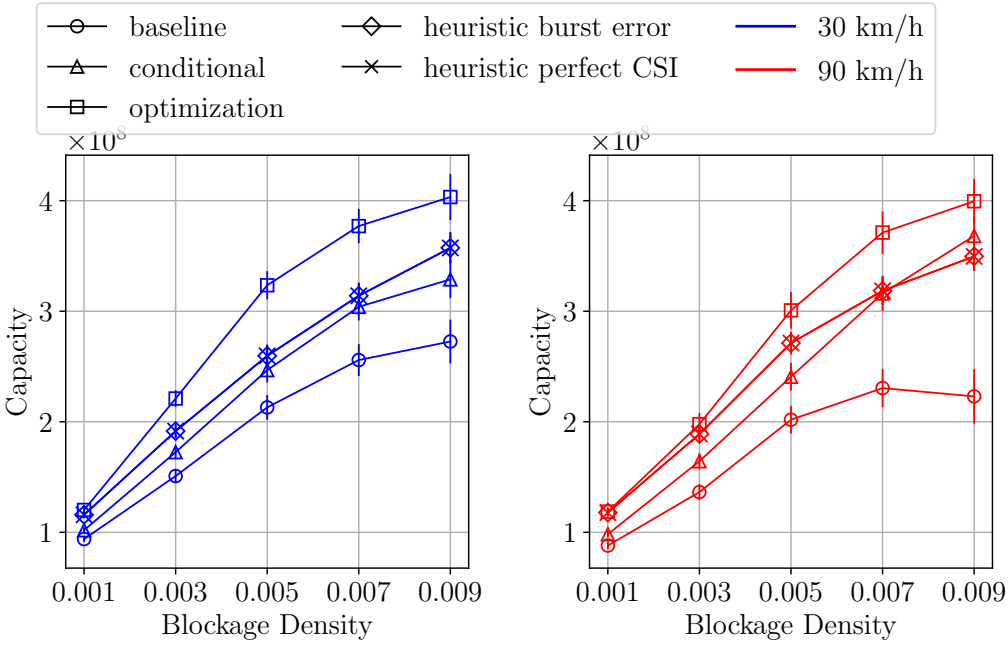
The capacity presents the same pattern as the SINR, with the optimization outperforming the other handover methods. In the same conditions as the other mechanisms, the baseline does not attain a comparable performance due to the poor channel conditions led by the handover choices made. As a consequence of the inferior propagation conditions, the theoretical capacity that the baseline can achieve is lower than the conditional, heuristic, and the optimization, as depicted in Fig. 4.13.

The quality of the channel conditions found by the UE is ultimately determined by the time the UE stays under outage conditions. The more time the UE is in outage the more the other KPIs average will drop. Figure 4.14 depicts the percentual time under outage for different values of blockage density. The baseline exhibits an outage increase when the velocity also increases and the blockage density is greater than 0.005, instead of decreasing like the other methods. This is reflected in the SNR, SINR, and capacity, as the baseline shows a faster decrease in the SNR and a slower increase in the SINR and capacity when compared to the other handover methods in the same conditions. This is due to the time under outage that increases when the blockage density and velocity also increase, as depicted in Fig. 4.14.



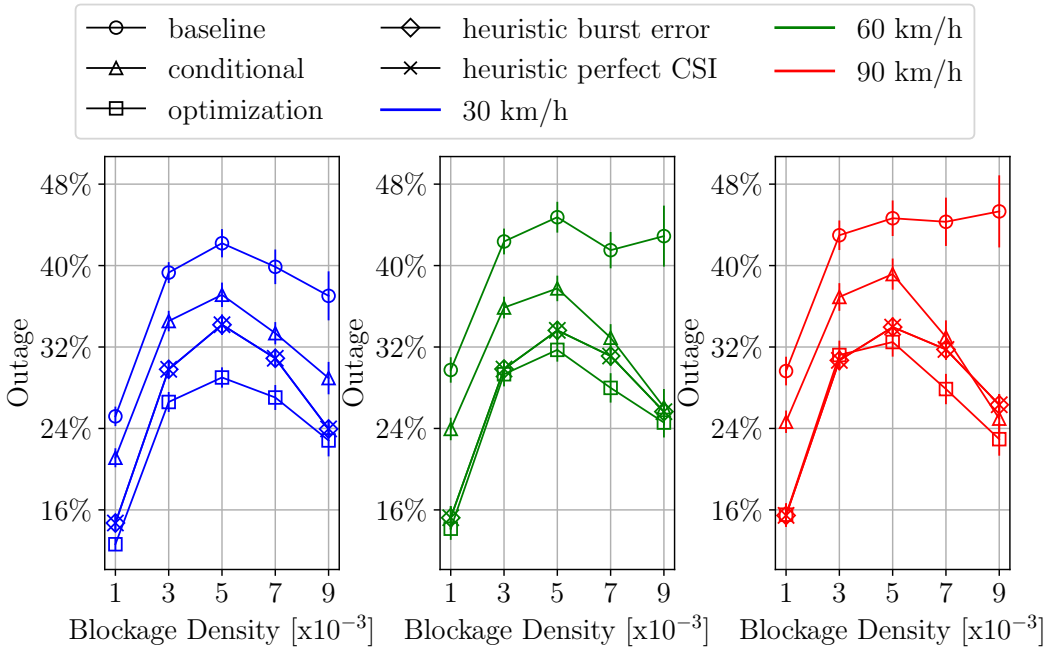
(a) 30 km/h (b) 90 km/h

Figure 4.12: SNR vs. blockage density



(a) 30 km/h (b) 90 km/h

Figure 4.13: Average Shannon capacity vs. blockage density



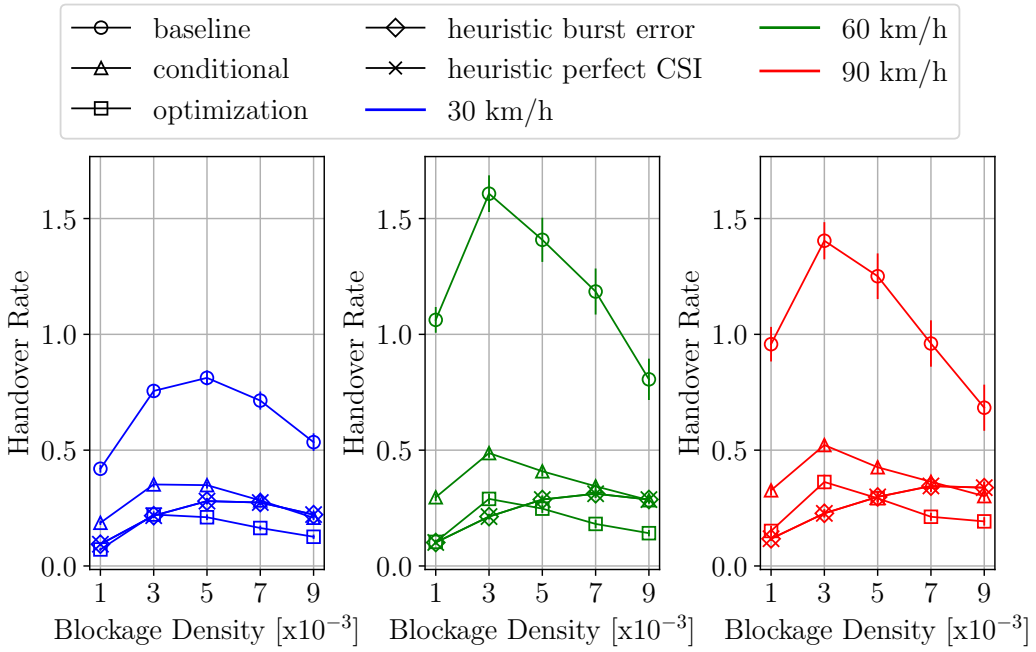
(a) 30 km/h (b) 60 km/h (c) 90 km/h

Figure 4.14: Percentual time under outage vs blockage density.

The conditional handover is better than the baseline handover in most scenarios. Because of the dual conditions, it is more likely that the condition handover will select a target BS that has a non-decreasing RSRP trend, as the preparation offset is usually below the execution offset. Thus, if the BS meets the preparation condition and the execution condition, this target BS is a better candidate than a BS that only fulfils the unique execution condition of the baseline. Therefore, with the conditional handover, the UE finds itself in better channel conditions to have a higher capacity as shown in Fig. 4.13.

Also, the heuristic handover method gap to the upper bound increases for scenarios with 90 km/h and blockage density greater than 0.007. Meanwhile, the conditional handover method gets closer to the MILP upper bound. This reveals that in the scenario of high blockage density and high velocity, the heuristic handover is not able to make good handover decisions from the blockage predictions. In this setting, the received powers from the TBSs are likely to be levelled due to the high probability of blockage. Thus, the heuristic method struggles to choose between TBSs to hand over and often makes the wrong decision or schedules many handovers in the same predictions. The conditional handover has more success in deciding which BS to hand over to due to the double condition that can capture whether a TBS has a non-decreasing RSRP.

In Fig. 4.15, we show the average handover rate, in handovers/s. First, the rate of handovers increases with the velocity, as expected. On the other hand, the

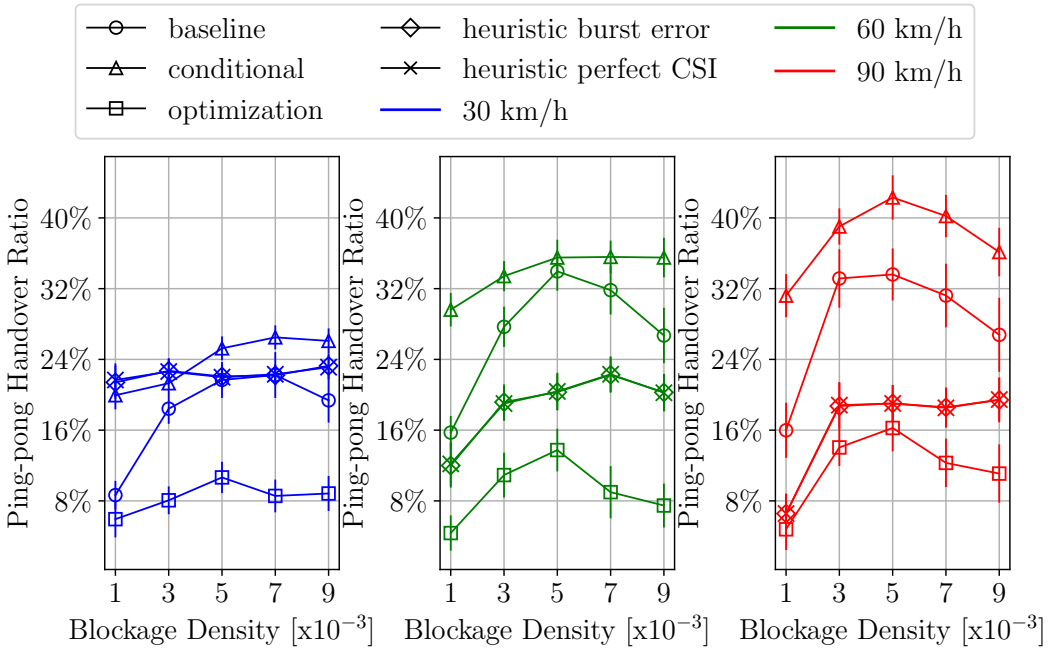


(a) 30 km/h (b) 60 km/h (c) 90 km/h

Figure 4.15: Average handover rate vs. blockage density.

number of handovers shows a non-monotonic trend with the increase in blockage density. Indeed, the number of handovers increases for lower blockage densities (0.001 obstacles/m^2) to compensate for the effect of more frequent and longer periods of blockage. However, as the blockage density increases beyond a certain point, the number of handovers starts decreasing. Such an effect occurs due to the channel deterioration that the UE faces when the blockage density is too high and the majority of the BSs are blocked. In other words, when the UE struggles with poor channel conditions due to severe blockage, this culminates in fewer target BS candidates one can hand over to.

One can notice that at 60 km/h, the baseline does more handovers than at 90 km/h, whereas the other handover mechanisms keep the increasing trend with velocity. We believe this is caused by a mismatch in the TTT or margin offset with the baseline at 60 km/h. In addition, Fig 4.15 shows that the optimization, the heuristic, and the conditional handover do fewer handovers than the baseline for the same velocity, approximately up to 70%, 67%, and 50% fewer handovers, respectively. Thanks to the thorough knowledge of channel conditions, the optimization makes better handover decisions than the baseline, which in contrast makes shortsighted decisions. We can confirm that the baseline handover choices are inefficient when looking at the ping-pong handover rate, shown in Fig. 4.16. The optimization reduced 50% of the ping-pong handover rate on average in comparison with the baseline. The large percentage of ping-pongs indicates that the 5G NR baseline



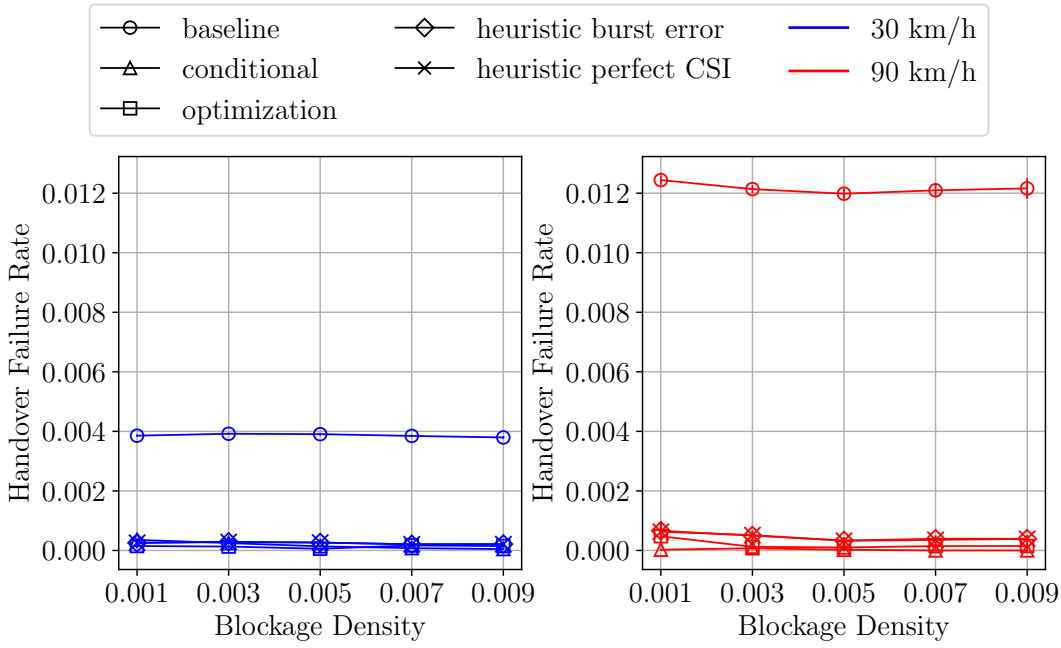
(a) 30 km/h (b) 60 km/h (c) 90 km/h

Figure 4.16: ping-pong handover ratio vs. blockage density.

handover decisions, only based on the instantaneous RSRP, are strongly affected by the blockage occurrences. Then, the handovers are mostly caused by the blockages, and rarely by the UE reaching the cell edge. Accordingly, such a large number of handovers increases the signalling overhead, BS load, and service interruption due to mobility.

The heuristic handover non-decreasing behaviour depicted in Fig. 4.15 for the increasing velocity scenario evidences what we described in Fig. 4.14. Rather than decreasing the handover rate for $\lambda \geq 0.005$, the heuristic handover method increases it, showing that it is erratic when the blockage density is high. To maximize the utility function, the algorithm tries to compensate for the blockage episodes with more handovers each time it receives a prediction and has to make a handover decision. Indeed, the heuristic handover ratio of ping-pongs also differs from the other methods, showing a non-decreasing trend for higher values of blockage densities. As the set of candidate BSs generating blockage prediction is distance-limited the dynamic programming algorithm likely induces these ping-pong handovers attempting to maximize the utility function value. Hence, the handover rate and the number of ping-pong handovers increase also harming other KPIs, such as outage episodes.

Although the conditional handover rate is lower than the baseline the ratio of ping-pong handovers is greater. Due to the increased reliability, the conditional handover succeeds in more handover attempts than the baseline handover. Proportionally, the baseline fails in more handovers than the conditional, including the



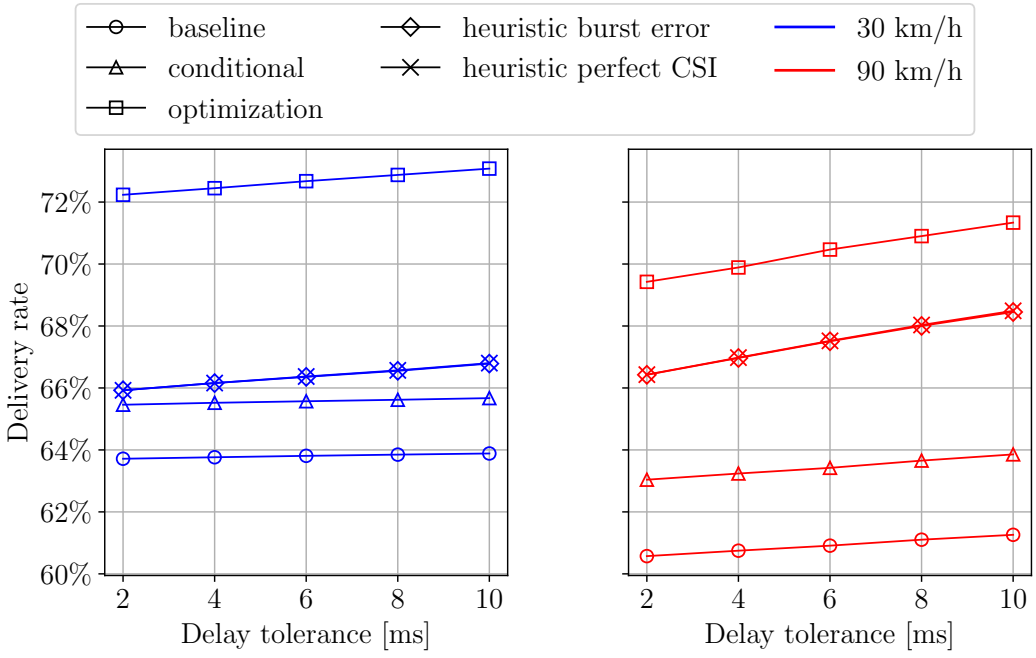
(a) 30 km/h (b) 90 km/h

Figure 4.17: Average handover failures per second. vs. blockage density.

ping-pong handovers. Fig. 4.17 depicts the average handover failure rate for the three mechanisms. The baseline outnumbers the other methods by a large margin, unveiling its weaknesses in the proposed scenario. Such a high rate of failures per second jeopardizes the UE performance due to the required connection re-establishment process that has to be carried out after a handover failure and lasts up to several seconds without a data connection at the UE.

In fact, the time the baseline is under outage is larger than the conditional, heuristic and optimization. Not only are the handover failures pushing the outages, but bad handover decisions also lead the baseline to BSs with poor channel conditions and massive handover attempts. On the other hand, conditional and the proposed heuristic handover methods significantly reduce the handover failures, so the differences in the optimization model are the handover decisions and the accumulated HIT. However, we consider the HIT a marginal contribution to the final outage, where RLFs and intervals of intense blockage and interference are the key factors impacting this performance.

The outage intervals jeopardize the baseline handover protocol performance by increasing the average message transmission delay, as illustrated in Fig. 4.18. The main sources of outages are connection re-establishment, handover failures, SINR below the out-of-sync threshold Q_{out}, and the handover-caused HIT. When the required delay is small, the baseline handover protocol struggles to send the messages before the latency deadline. On the other hand, the optimization shows a constant



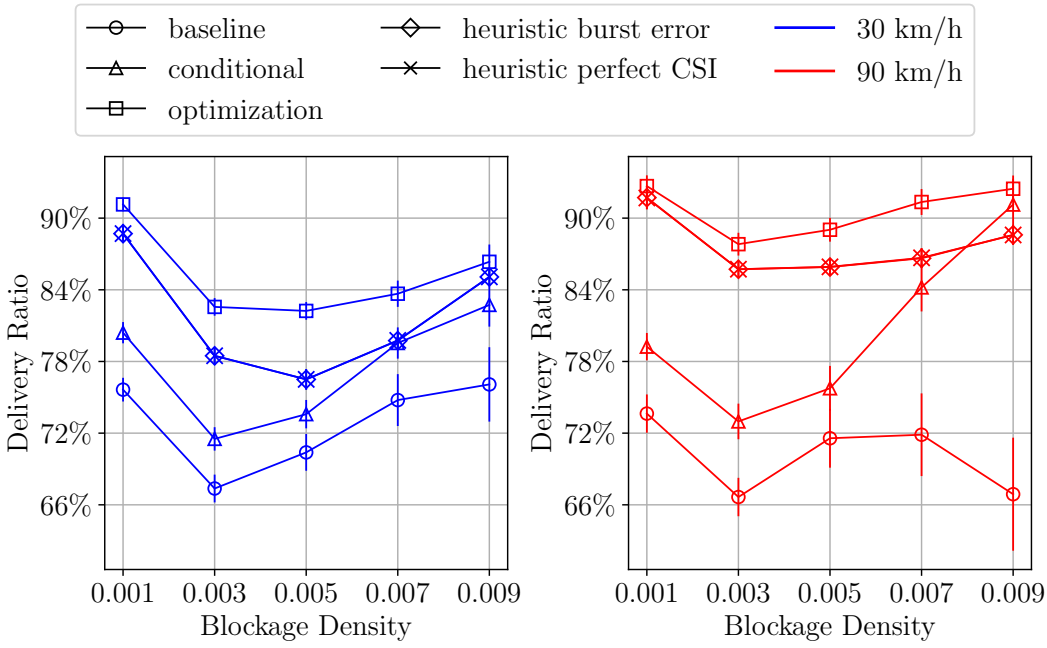
(a) 30 km/h (b) 90 km/h

Figure 4.18: Packet delivery rate meeting the constraint set by the delay tolerance.

packet delivery rate despite the tight delay requirement. For instance, the baseline sent 12% fewer messages out of the delay window for both, the low and high speeds, than the optimization.

The delivery rate in Fig. 4.18 is too low because of the strict delay tolerance required by URLLC applications, less than 10 ms. Figure 4.19 depicts the average delivery rate by the blockage density. For the lower values of blockage density, the baseline and conditional handover deliver up to 15% and 10% fewer packets than the optimization does. This gap decreases as the blockage density increases, except for the baseline at 90 km/h due to the intense outage. Again, the heuristic handover technique presents the performance closest to the upper bound for almost all the blockage density values, having 6% as the largest gap to the MILP-provided upper bound. The conditional handover is more effective than the heuristic solely in the blockage density equal to 0.009 and velocity equal to 90 km/h. In such a situation, the heuristic handover has an increased handover rate, which might impede the packet transmissions due to the HIT of the multiple handovers or the lack of BS with good channel conditions in range for blockage prediction.

Regarding the blockage prediction error function, one may argue that the metrics survey up to this point did not show how it would have impacted the dynamic programming algorithm. In fact, the error function degrades the blockage prediction but it marginally affects the overall handover performance. We cannot state that the heuristic handover is robust to error, but the error functions tested in the



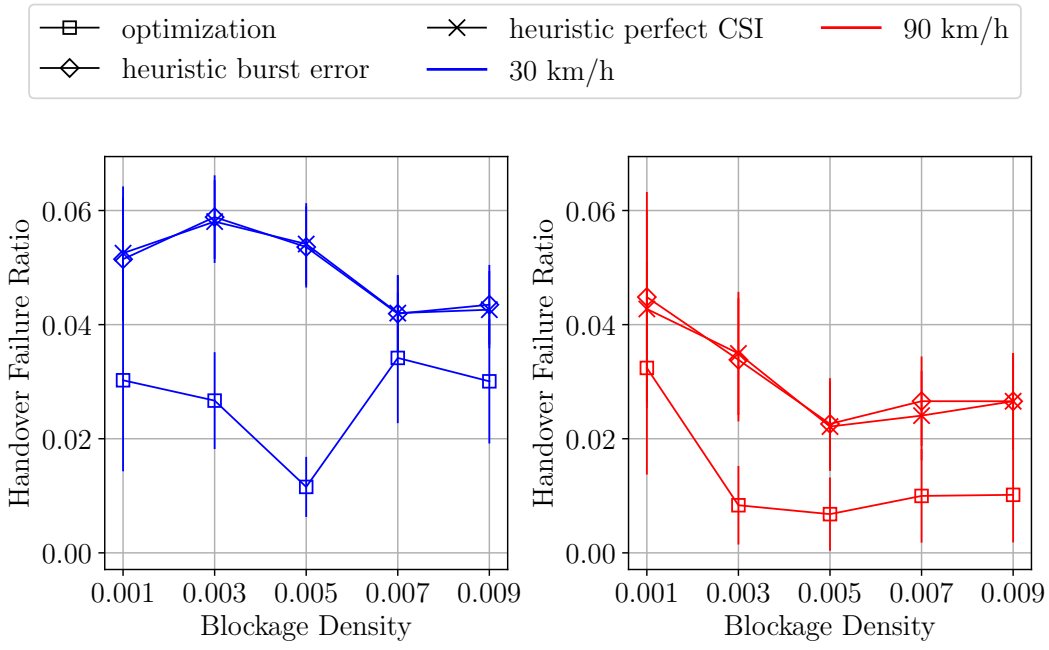
(a) 30 km/h (b) 90 km/h

Figure 4.19: Delivery rate meeting the constraint set by the delay tolerance.

preliminary evaluations of this thesis showed that the chosen one was the most severe and its impact was still minor. Among the error functions tested, we used a Normal distribution to induce error in the beginnings and endings of the blockage episodes with a standard deviation that increases with the prediction slot, and also a Poisson-distributed burst length with a linear error function.

We infer that the limited set of BSs available to handover and the constraint that these BSs must be inside a distance range results in better channel conditions than handover to distant BSs. Also, the augmented handover rate in high-blockage density scenarios exposes this tendency to make more handovers to cope with hampering channel conditions. We highlight the difference between the performance with perfect prediction and the prediction with errors, as shown in Fig. 4.20. In this Figure, we can notice that the heuristic handover has a similar performance with errors and perfect blockage predictions. It is easy to see that the average points and the confidence intervals overlap for all blockage densities. For higher velocities, an opposite trend in relation to the SINR is also noticeable. The SINR is the criteria to decide whether a handover is successful either due to a failed message exchange between SBS and UE or an unsuccessful RACH. Finding a proper error function to emulate the blockage prediction inaccuracies and testing the heuristic handover algorithm's robustness to prediction error has high priority in our future works.

The conditional handover must be particularly assessed in terms of preparation and release ratio. Both metrics highlight the efficiency of the conditional handover

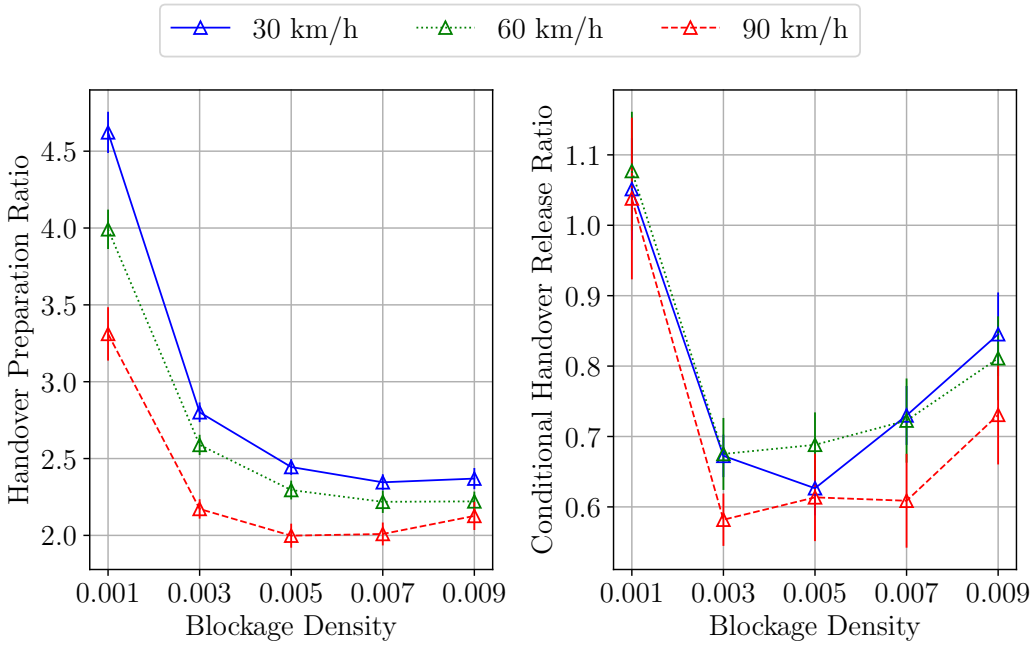


(a) 30 km/h (b) 90 km/h

Figure 4.20: Average handover failures per second vs. blockage density detailed for the heuristic handover with blockage prediction error, without error and the MILP formulation.

parameters configuration and the overhead generated by the conditional mechanism. First, the preparations are more numerous when the blockage density is lower because more BS are in LoS condition being capable of reaching the preparation offset, as evidenced by Fig. 4.21a. The drawback is that each preparation implies a handover request negotiation between the SBS and the candidate TBS. Thus, if this number of preparations is too high, the overhead may escalate the network congestion. For example, for λ equal to 0.001 obstacles/m^2 there are almost 5 preparations for each confirmed handover, while that is about 3 for 0.003 at 30 km/h.

Similarly, the releases are prejudicial in terms of overhead, but they avoid a BS in poor channel conditions being considered to hand over and occupy a spot in the execution list. There is a decrease in the releases as blockage density increases, as depicted in Fig. 4.21b. However, there is a turning point as the blockage density increases, specifically for λ greater or equal to 0.005. For such values of blockage density, the average RSRP, and thus more BSs, are found in the release condition than for intermediate values of λ. Likewise in the handover rate and ping-pong handover ratio, this is the case where for low values of blockage density, the SBS is better than the other BSs because it is in LoS and the channel conditions are too good, as evidenced by the SNR, until the NLoS probability rises, making the BSs' channel get worse and only the closest ones stay above the release threshold. A handover parameters configuration algorithm which adjusts the conditional han-



(a) Conditional handover preparation ratio (b) Conditional handover release ratio

Figure 4.21: Conditional handover preparation ratio vs blockage density, and Conditional handover release ratio vs blockage density.

dover parameters to the surrounding situation of the scenario might be a solution to this challenge. For example, using the blockage prediction to infer the blockage density in the surroundings and adjust the conditions handover parameters to avoid excessive handover preparations.

Regarding the performance handicaps due to excessive handovers, the literature suggests reducing the context exchanging time and HIT, using dual-connection [51] and improving coverage [13]. Likewise, the results in this thesis point out that efficient handover policies to avoid ping-pongs and failures might also enhance the protocol performance, by reducing the control-plane overhead due to handover control messages and achieving high throughput and low latency. The conditional handover outperforms the baseline, presenting robustness to channel variations, confirmed by the fewer handover failures and fewer outages. Although the handover decisions are not the best ones leading to the higher ping-pong rate, they are still better than the baseline handover decisions due to the double condition.

The optimization model, however, needs to have all the UE data prior to the optimization model solving, which rather than for real-time handover applications, makes the proposed model more relevant for offline applications, such as benchmarking of handover mechanisms and network planning for mobile and vehicular networks. Furthermore, the proposed optimization tool can prove to be a very useful tool for evaluating the strengths and weaknesses and indicating areas of improvement of currently adopted handover mechanisms. The heuristic handover method is

feasible for real-time applications. The heuristic handover is based on short blockage predictions and a limited set of BSs in range. The heuristic handover attains a performance close to the optimization upper bound in most metrics evaluated. However, the drawback of deploying the heuristic handover technique is the infrastructure and computational power to generate blockage predictions, who is already considered enablers of future mobile communication technologies.

4.5 Conclusion

In this thesis, we propose a joint handover and user performance-aware MILP optimization mechanism for mmWave networks with a focus on high-capacity URLLC 5G NR networks, such as vehicular networks. The preliminary results with this MILP formulation motivated the development of a heuristic with lower complexity with the same concepts, i.e., maximizing a utility function based on future information. We propose a dynamic programming heuristic that iterates over predicted CSI generated from blockage predictions to maximize a utility function based on the original optimization problem.

When compared with the standardized 5G NR baseline and conditional handover mechanisms, the proposed optimization is more than 50% less prone to the ping-pong effect, on average. The conditional handover shows robustness to outage and handover failure with a maximum 40% outage against 49% of the baseline. Additionally, the handover decisions taken by the proposed optimization and heuristic approaches lead to an average decrease in the handover rate of up to 71% and 67%, respectively, hence reducing the impact of HIT and avoiding packets being discarded due to excessive latency.

Chapter 5

Conclusion

This chapter discusses the thesis outcomes and our ideas for future works. In this thesis, we detailed the 3GPP 5G baseline and conditional handover techniques, presented a novel MILP formulation for handover optimization and developed a heuristic handover based on blockage prediction to improve handover decision-making in the context of 5G NR mmWave URLLC. We carried out a numeric evaluation that led to valuable insights about these handover mechanisms, which we discuss in this chapter highlighting their weaknesses and strengths. Finally, we propose future work opportunities to expand our heuristic handover investigations and also approach handover optimization using federated machine learning.

5.1 Final Remarks

The fifth generation of mobile communication (5G) evolved from the previous generation with the 3GPP intention to provide massive data rate connection everywhere, anytime and with energy efficiency. Recent advancements in signal processing technology such as Massive MIMO (M-MIMO) and Network Functions Virtualization (NFV) pushed these ambitious goals to become reality. However, mobility is still an open problem, particularly Vehicle-to-Everything (V2X) applications that might enable unmanned driving and provide many benefits, such as road security.

Some V2X use cases have strict capacity, delay and reliability requirements. Implementing these use cases in the sub-6 GHz band is challenging due to its high co-channel interference that limits the achievable capacity. Then, the mmWave emerges as a solution but imposes the directionality and propagation issues. These conditions make mobility more challenging in the mmWave, thus efficient mobility-handling, such as handovers, are crucial enablers of autonomous vehicles, beyond field of view UAV, and other V2X applications.

The 3GPP proposed two handover algorithms in the 5G NR context, the baseline and the conditional handover, detailed in Chapter 2. The baseline handover simply

compares the periodically advertised RSRP of the surrounding BSs with the SBS and initiates a handover if this condition is true. The conditional handover uses two conditions instead of only one and decouples the handover preparation from the handover execution. These two features allow the conditional handover mechanism (i) to select a BS whose RSRP is non-decreasing, once the preparation condition RSRP margin is lower than the execution condition RSRP margin, and (ii) the user can continue communicating with the SBS while waiting the execution to the TBS without depending on the SBS for any handover signalling.

In this thesis, we approached handover optimization with a novel MILP framework for handover optimization and a heuristic handover method, both presented in Chapter 3. The MILP formulation jointly optimizes the handover and the KPIs of interest. We propose three objective functions, aggregated capacity maximization, joint capacity maximization and delay assurance, and joint capacity maximization and outage minimization. This optimization problem is too complex to be solved in a combinatorial fashion, as proved in Appendix A, thus the MILP formulation offers the solution with a lower computational cost. However, the mathematical programming formulation requires all the channel information, from all the BSs and for all timeslots beforehand to come up with a solution, which is an unrealistic premise. Therefore, this optimization framework design privileges evaluating other handover mechanisms and network configurations.

Though the MILP formulation is not a valid option for a real-time implementation we have devised a solution based on the optimization's strengths. Although we can not assume that all the CSI will be available at the beginning of the process, it is possible to predict the channel status a few seconds ahead with considerable accuracy [17]. Additionally, we can reduce the search space by restricting the number of BSs considered candidates for a handover, including those inside a given range. With a dynamic programming approach and taking these two assumptions, the heuristic handover is computationally more efficient and quickly responsive for a handover decision-making algorithm.

Implementing the heuristic imposes an infrastructure challenge to provide the blockage predictions. The BSs must have LIDAR sensors and periodically send the sensor scans to a CPU. The CPU is the entity that processes the computational load, split into two main activities: generate the blockage predictions and turn them into a handover schedule. We recognize CPU scalability as a problem, but this is per se a problem to be tackled and studied in many areas, such as Cell-Free Massive MIMO (CF MMIMO) [92]. Besides, we assumed modifications in the baseline handover protocol but exploited the same handover control messages and phases. The O-RAN alliance [101] envisions context-base handovers as a use-case for Near Real-Time RAN Intelligent Controller (Near RT-RIC) [102], thus giving

the possibility to realize the modifications in the handover protocol [93].

We have implemented the simulation code for all the handover methods in Python. All the simulations received as input the same instance, containing a scenario description with the position and resources of the BSs, a description of the UE position and details of the mobility model, the obstacles positions, and the blockage status of each UE-BS pair for each timeslot. This instance is pre-processed to generate the scenario and deploy the UE and BSs, in the case of the baseline and conditional handover simulations, and also generates the CSI beforehand for the optimization and heuristic handover. Each simulation then follows its main routine to generate the association through handovers. With the associations, the post-processing calculates the KPI that will be used to assess the handover techniques.

The baseline and conditional handover are implementations of a handover interface and all the code for their simulation is implemented using interfaces that can be easily extended. A discrete event simulation library applies the 5G NR UE life-cycle and executes handover using the baseline and conditional classes when necessary. The optimization is implemented using the Gurobi Optimizer solver's Python library, so the main routine creates the model and adds the variables, constraint, and objective function. Once the model is complete, the solver analyzes the model and applies mathematical programming techniques to solve it, storing the result in the variables that are then transformed into the UE association. Finally, the heuristic uses the optimization pre-processing, but the main function iterates for each timeslot updating the UE position and the periodical blockage predictions that generate the handover schedule.

The three handover mechanisms, the baseline and conditional 3GPP handover, and the heuristic handover were assessed in Chapter 4 and compared with the MILP-provided upper bound. Our evaluation begins by surveying the handover's parameters Time to Trigger (TTT) and the offset margin. Our findings point out that the TTT must be adjusted to UE velocity to respond according to the approaching/distancing of the cell border without allowing unnecessary handover or denying the necessary ones. Regarding the margin, the baseline handover is the most affected by the margin decreasing which increases the handover rate and the ratio of ping-pongs, though increasing the margin might also increase handover failures.

In terms of velocity and blockage density, the heuristic handover presents the closest performance to the optimization. Concerning delivery rate the heuristic is at most 6% worse than the optimization and almost 4% in outage. However, we found that for higher blockage density values the heuristic method ramps up the number of handovers to cope with the unstable channel condition generated by blockage. In such scenarios, the conditional and heuristic handover have comparable performance regarding capacity, period under outage, and packet delivery rate. The

conditional handover is, in turn, more efficient than the baseline in most metrics, such as robustness to handover failure, capacity, and outage, but still has flaws regarding the handover decisions. Such flaws are evident in the higher number of ping-pong handovers and also the elevated number of preparations in low blockage density scenarios.

Finally, we consider it useful to have future information through prediction, giving a considerable advantage for the handover mechanisms. It is a concern that requesting and disseminating such information may have the undesired effect of increasing overhead. Therefore, this information must be delivered on time and accurately, to avoid misled handover decisions based on stale predictions and with a high overhead cost. Besides the prediction information, whichever algorithm generates the handover decision must have the means to do it quickly, for example, by limiting the prediction length or the set of BS.

5.2 Future Works

Although the results in Chapter 4 justify the importance of the optimization model, which successfully outperformed the other handover mechanisms in terms of handover ping-pong effect, capacity, and delay, it is not practical to assume that it will be implemented for making real-time handover decisions. First, assuming that the UE or BS will know all the further blockage events for a relatively long period ahead is unreasonable. Second, the time to compute the model solution is considerably longer than the period available to make a handover decision in a URLLC system.

However, combining the optimization model and baseline constitutes a powerful framework for handover optimization evaluation. With these bounds, we can assess how far a handover optimization strategy is with respect to the optimization model and the baseline. Thus, in our future works, we will concentrate on continuing the investigation of the heuristic handover and extend our research to approach ML.

5.2.1 Scenario

In this thesis, we tested a scenario with a single vehicle roaming through a straight road. We reckon that our evaluation would benefit from a more diverse scenario. However, due to the evaluation complexity and the thesis timeline, we could not implement a different scenario in the current version. With the current results validated, we can assess the handover algorithms in different scenarios and distinguish the scenario impacts from the handover algorithms' performance effect on the metrics. It is worth mentioning that the codes generated can already run different input instances, as soon as they comply with the expected format.

Similarly, we intend to expand our evaluation to multiple users. That would bring the resource allocation aspect to this work, and so it would need careful load balancing and a user schedule scheme. Again, it was important in this phase of the work to isolate and differentiate the influence of handover algorithms on UE's experienced performance. In a future version that contemplates multiple vehicular users, the handover admission control and the timeslot allocation will play a key role in a simulated scenario.

5.2.2 Heuristic

The heuristic introduced in this thesis achieved a satisfactory performance compared to the baseline and conditional handover but still, there is some gap to the upper bound. In high blockage density values, the heuristic handover presented a high handover rate and increased the ratio of ping-pong handovers instead of decreasing such as the other handover mechanisms. Albeit its performance results, we also want to improve our evaluation and understand how the heuristic would behave in different scenarios.

Assessing a mechanism based on prediction motivates the numerical evaluation of the mechanism's robustness to the prediction error. When testing the heuristic handover algorithm, we tried to match the error functions in the blockage prediction literature. However, in the timeframe of the thesis, we could not find an error function that led to a significant difference between perfect blockage prediction and different levels of errors. This can result from the small distance range to include a BS as a candidate to handover, which limited the choices but conversely kept a lower path loss attenuation to the BSs.

Likewise, there are other factors to be evaluated. First, the impact of the range to include the BSs that might directly affect the handover rate and ping-pong ratio. Besides, as aforementioned, we believe a larger range could make the heuristic handover mechanism more sensible to prediction errors. Consequently, the handover rate, ping-pong handovers, and failures ratio would increase and exacerbate the effect of high blockage densities on the heuristic. Decreasing the range would limit the options of BSs, reducing the number of handovers below the necessary to keep the performance acceptable. Therefore, there might exist a trade-off between the radius of BSs considered by the heuristic and the prediction error impact on heuristic handover performance which is a valuable opportunity for further investigations.

Another possible parameter to be evaluated is the number of BSs that can generate predictions. It is reasonable to assume that not all the BSs will embed a LIDAR sensor. Because of a gradual upgrade of the RAN infrastructure or CAPEX limitations, blockage prediction-capable and non-capable BSs might coexist in the

same scenario. Thus the algorithm must be modified to this condition and make handover decisions with blockage prediction data and RSRP only. Having a partial blockage prediction coverage could considerably impact the quality of the handover decisions made by the heuristic handover and new strategies are needed to overcome this, such as inferring the future CSI of the BSs without LIDAR sensor using the history of RSRP and the blockage prediction available.

We mentioned earlier in Chapter 4 that it was challenging to implement the heuristic as a handover class like the baseline and conditional handover. Such modification requires significant changes to the stable version of the UE and BS objects implementation used in the baseline and conditional handover simulations to support the scheduling of blockage predictions. Thus, we decided to postpone this modification and followed the numerical evaluations with a different simulation code. Another interesting possibility in the future is the opportunity to apply the heuristic in the O-RAN Near RT-RIC as a proof of concept closer to a realistic scenario.

Blockage prediction is not the only information available that can be used to develop an efficient handover algorithm. Although blockage prediction had a high synergy with our framework, other information such as UE position, velocity, and obstacles' position, are available in our scenario and could be used along with blockage prediction. New heuristics or a different version of our heuristic tailored to these other available data might improve the performance and approximate even more to the optimization upper bound.

5.2.3 Machine Learning Framework

Recently, ML has gained a lot of momentum and become a consolidated trend in the academic and industrial communities. Owing to the generalization capabilities the ML algorithms present, many intractable problems gained easier solutions through it. In wireless networks, for example, ML has numerous applications, such as received power prediction, parameter tuning, resource allocation, and channel decoding [103]. However, the ML algorithms have many parameters to set, and most of the relevant works in the literature *do not*: (*i*) develop upper bounds to the problem, (*ii*) evaluate the algorithm's effectiveness for the problem, and (*iii*) weigh the impact on privacy and security.

Another thing to be concerned with is the features chosen to compose the ML model. In a problem such as the handover optimization, many factors may influence the handover decision, from the handover parameters, and UE velocity, to the blockage predictions and past CSI. Thus, the features must improve the ML model towards optimization and avoid overfitting. The desired features will present a low correlation with the other features and a high correlation with the positive han-

dover decisions. The optimization can serve as a ground-truth classifier that labels the handover decisions, training a supervised learning model to accept or deny the handovers.

In collaboration with Prof Indrakish Dei, from South East Technological University (SETU), we designed an ML framework for the handover optimization problem. This model is based on two main building blocks, associative rule mining and federated learning. The two building blocks also define two stages of the framework operation. In the earliest stage, associative rule mining techniques are applied to the dataset to find an initial model. This initial model will learn from the KPI and side information available, such as RSRP, Delay, UE position, and velocity, to which BS handover and when. In the second stage, the model is propagated in the network to allow initial operation. In the last step, the model is constantly and locally updated by the users and fed back to the network, which in turn averages the collection of data from many users and then updates the model globally.

The associative rule mining gives a custom initial model, already suited for our problem, and the model is successively refined by federated learning. This approach has challenges to be tackled. First, the framework relies on intense communication between the network and the UE, which can be expensive [104], mainly in the mmWave band. Second, in a heterogeneous mobility scenario, handover-assisting data can change from one area to another, so one might need to rely on local models instead of a central model.

References

[1] 3GPP. *5G; NR; NR and NG-RAN Overall description.* Technical Specification (TS) 38.300, 3rd Generation Partnership Project (3GPP), 07 2020. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138300_138399/138300/16.02.00_60/ts_138300v160200p.pdf>. Version 16.2.0.

[2] 3GPP. *5G; NR; Base Station (BS) radio transmission and reception.* Technical Specification (TS) 38.104, 3rd Generation Partnership Project (3GPP), 07 2020. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138100_138199/138104/16.04.00_60/ts_138104v160400p.pdf>. Version 16.4.0.

[3] 3GPP. *5G; NR; Radio Resource Control (RRC); Protocol specification.* Technical Specification (TS) 38.331, 3rd Generation Partnership Project (3GPP), 10 2018. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138300_138399/138331/15.03.00_60/ts_138331v150300p.pdf>. Version 15.3.0.

[4] KANG, S., CHOI, S., LEE, G., et al. "A dual-connection based handover scheme for ultra-dense millimeter-wave cellular networks". In: *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. IEEE, 2019.

[5] 3GPP. *Evaluation and design aspects for NR mobility enhancement.* Technical Document (TDoc) 168852, 3rd Generation Partnership Project (3GPP), 11 2016. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138300_138399/138331/15.03.00_60/ts_138331v150300p.pdf>. R2.

[6] ARAUJO, G., ALERIGI, A. "Telecom giants shell out $1.27 bln in first day of Brazil's 5G auction", *Reuters*, 2021. Disponível em: <https://www.reuters.com/business/media-telecom/brazil-holds-long-awaited-5g-spectrum-auction-eyeing-8-bln-investments-2021-11-04/>.

[7] ALERIGI, A., BAUTZER, T. "Brazil to reschedule auction for unsold 5G spectrum, minister says", *Reuters*, 2021. Disponível em: <https://www.reuters.com/business/media-telecom/brazil-reschedule-auction-unsold-5g-spectrum-minister-says-2021-11-05/>.

[8] POCOVI, G., SHARIATMADARI, H., BERARDINELLI, G., et al. "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements", *IEEE Network*, v. 32, n. 2, pp. 8–15, 2018.

[9] FORD, R., ZHANG, M., MEZZAVILLA, M., et al. "Achieving Ultra-Low Latency in 5G Millimeter Wave Cellular Networks", *IEEE Communications Magazine*, v. 55, n. 3, pp. 196–203, 2017. doi: {10.1109/MCOM.2017.1600407CM}.

[10] KONG, L., KHAN, M. K., WU, F., et al. "Millimeter-wave wireless communications for IoT-cloud supported autonomous vehicles: Overview, design, and challenges", *IEEE Communications Magazine*, v. 55, n. 1, pp. 62–68, 2017.

[11] LIEN, S.-Y., KUO, Y.-C., DENG, D.-J., et al. "Latency-optimal mmWave radio access for V2X supporting next generation driving use cases", *IEEE Access*, v. 7, pp. 6782–6795, 2018.

[12] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification.* Technical Specification (TS) 36.331, 3rd Generation Partnership Project (3GPP), 06 2024. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/36_series/36.331/36331-i20.zip>. Version 18.2.0.

[13] TAYYAB, M., GELABERT, X., JÄNTTI, R. "A survey on handover management: From LTE to NR", *IEEE Access*, v. 7, pp. 118907–118930, 2019.

[14] USMAN, M. A., PHILIP, N. Y., POLITIS, C. "5G enabled mobile healthcare for ambulances". In: *2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6. IEEE, 2019.

[15] HAGHRAH, A., ABDOLLAHI, M. P., AZARHAVA, H., et al. "A survey on the handover management in 5G-NR cellular networks: aspects, approaches and challenges", *EURASIP Journal on Wireless Communications and Networking*, v. 2023, n. 1, pp. 52, 2023.

[16] SUN, L., HOU, J., SHU, T. "Optimal handover policy for mmWave cellular networks: A multi-armed bandit approach". In: *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. IEEE, 2019.

[17] WU, S., CHAKRABARTI, C., ALKHATEEB, A. "Proactively Predicting Dynamic 6G Link Blockages Using LiDAR and In-Band Signatures", *IEEE Open Journal of the Communications Society*, v. 4, pp. 392–412, 2023. doi: 10.1109/OJCOMS.2023.3239434.

[18] BOER, A., VAN DE VELDE, R., DE VRIES, M. "Self-driving vehicles (SDVs) & geo-information". 2017.

[19] GIMENEZ, L. C., MICHAELSEN, P. H., PEDERSEN, K. I., et al. "Towards zero data interruption time with enhanced synchronous handover". In: *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–6. IEEE, 2017.

[20] MBULWA, A. I., YEW, H. T., CHEKIMA, A., et al. "Self-Optimization of Handover Control Parameters for 5G Wireless Networks and Beyond", *IEEE Access*, v. 12, pp. 6117–6135, 2024. doi: 10.1109/ACCESS.2023.3346039.

[21] QI, K., LIU, T., YANG, C., et al. "Dual connectivity-aided proactive handover and resource reservation for mobile users", *IEEE Access*, v. 9, pp. 36100–36113, 2021.

[22] LEE, C., JUNG, J., CHUNG, J.-M. "Intelligent dual active protocol stack handover based on double DQN deep reinforcement learning for 5G mmWave networks", *IEEE Transactions on Vehicular Technology*, v. 71, n. 7, pp. 7572–7584, 2022.

[23] WU, J., FAN, P. "A Survey on High Mobility Wireless Communications: Challenges, Opportunities and Solutions", *IEEE Access*, v. 4, pp. 450–476, 2016. doi: 10.1109/ACCESS.2016.2518085.

[24] POLESE, M., GIORDANI, M., MEZZAVILLA, M., et al. "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks", *IEEE Journal on Selected Areas in Communications*, v. 35, n. 9, pp. 2069–2084, 2017. doi: 10.1109/JSAC.2017.2720338.

[25] MASRI, A., VEIJALAINEN, T., MARTIKAINEN, H., et al. "Machine-learning-based predictive handover". In: *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pp. 648–652. IEEE, 2021.

[26] STANCZAK, J., KARABULUT, U., AWADA, A. "Conditional handover in 5G-principles, future use cases and FR2 performance". In: *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 660–665. IEEE, 2022.

[27] LEE, C., CHO, H., SONG, S., et al. "Prediction-based conditional handover for 5G mm-wave networks: A deep-learning approach", *IEEE Vehicular Technology Magazine*, v. 15, n. 1, pp. 54–62, 2020.

[28] ZAKRZEWSKA, A., D'ANDREAGIOVANNI, F., RUEPP, S., et al. "Biobjective optimization of radio access technology selection and resource allocation in heterogeneous wireless networks". In: *2013 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 652–658. IEEE, 2013.

[29] KHOSRAVI, S., SHOKRI-GHADIKOLAEI, H., PETROVA, M. "Learning-based handover in mobile millimeter-wave networks", *IEEE Transactions on Cognitive Communications and Networking*, v. 7, n. 2, pp. 663–674, 2020.

[30] WU, S., CHAKRABARTI, C., ALKHATEEB, A. "Lidar-aided mobile blockage prediction in real-world millimeter wave systems". In: *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2631–2636. IEEE, 2022.

[31] NISHIO, T., OKAMOTO, H., NAKASHIMA, K., et al. "Proactive received power prediction using machine learning and depth images for mmWave networks", *IEEE Journal on Selected Areas in Communications*, v. 37, n. 11, pp. 2413–2427, 2019.

[32] ANDREWS, J. G., BUZZI, S., CHOI, W., et al. "What will 5G be?" *IEEE Journal on selected areas in communications*, v. 32, n. 6, pp. 1065–1082, 2014.

[33] GIORDANI, M., ZANELLA, A., HIGUCHI, T., et al. "Performance study of LTE and mmWave in vehicle-to-network communications". In: *2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, pp. 1–7. IEEE, 2018.

[34] RAPPAPORT, T. S., SUN, S., MAYZUS, R., et al. "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, v. 1, pp. 335–349, 2013.

[35] PARK, J.-J., LEE, J., LIANG, J., et al. "Millimeter wave vehicular blockage characteristics based on 28 GHz measurements". In: *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5. IEEE, 2017.

[36] WANG, Q., ZHAO, X., LI, S., et al. "Attenuation by a human body and trees as well as material penetration loss in 26 and 39 GHz millimeter wave bands", *International Journal of Antennas and Propagation*, v. 2017, 2017.

[37] MOLISCH, A. F., RATNAM, V. V., HAN, S., et al. "Hybrid beamforming for massive MIMO: A survey", *IEEE Communications magazine*, v. 55, n. 9, pp. 134–141, 2017.

[38] MABROUKI, S., DAYOUB, I., LI, Q., et al. "Codebook Designs for Millimeter-Wave Communication Systems in Both Low-and High-Mobility: Achievements and Challenges", *IEEE Access*, v. 10, pp. 25786–25810, 2022.

[39] NIU, Y., LI, Y., JIN, D., et al. "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges", *Wireless networks*, v. 21, pp. 2657–2676, 2015.

[40] SORET, B., MOGENSEN, P., PEDERSEN, K. I., et al. "Fundamental tradeoffs among reliability, latency and throughput in cellular networks". In: *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1391–1396. IEEE, 2014.

[41] SCOTTSDALE, A. "635 New 5G Cities in 2021; 1,947 5G Cities Globally, According to VIAVI", *Viavi*, 2022. Disponível em: <https://www.viavisolutions.com/en-us/news-releases/635-new-5g-cities-2021-1947-5g-cities-globally-according-viavi>.

[42] ERICSSON. "Network coverage outlook", *Ericsson*, 2022. Disponível em: <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/network-coverage>.

[43] OKANO, M., HASEGAWA, Y., KANAI, K., et al. "Field experiments of 28 ghz band 5g system at indoor train station platform". In: *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1–6. IEEE, 2020.

[44] ZHAO, Y., WEI, M., HU, C., et al. "Latency analysis and field trial for 5g nr". In: *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5. IEEE, 2022.

[45] LIU, GUANGYI AND HOU, XEUYING AND HUANG, YOUHONG AND SHAO, HUA AND ZHENG, YI AND WANG, FEI AND WANG, QIXING. "Coverage enhancement and fundamental performance of 5G: Analysis and field trial", *IEEE Communications Magazine*, v. 57, n. 6, pp. 126–131, 2019.

[46] FRENGER, P., TANO, R. "More capacity and less power: How 5G NR can reduce network energy consumption". In: *2019 IEEE 89th vehicular technology conference (VTC2019-Spring)*, pp. 1–5. IEEE, 2019.

[47] LIN, X., LI, J., BALDEMAIR, R., et al. "5G new radio: Unveiling the essentials of the next generation wireless access technology", *IEEE Communications Standards Magazine*, v. 3, n. 3, pp. 30–37, 2019.

[48] GUPTA, A., JHA, R. K. "A survey of 5G network: Architecture and emerging technologies", *IEEE access*, v. 3, pp. 1206–1232, 2015.

[49] VASUDEVA, K., SIMSEK, M., LÓPEZ-PÉREZ, D., et al. "Analysis of handover failures in heterogeneous networks with fading", *IEEE Transactions on Vehicular Technology*, v. 66, n. 7, pp. 6060–6074, 2016.

[50] KARABULUT, U., AWADA, A., VIERING, I., et al. "RACH optimization with decision tree based supervised learning for conditional handover in 5G beamformed systems", *arXiv preprint arXiv:1910.11890*, 2019.

[51] PARK, H.-S., LEE, Y., KIM, T.-J., et al. "Handover mechanism in NR for ultra-reliable low-latency communications", *IEEE Network*, v. 32, n. 2, pp. 41–47, 2018.

[52] LAURIDSEN, M., GIMENEZ, L. C., RODRIGUEZ, I., et al. "From LTE to 5G for connected mobility", *IEEE Communications Magazine*, v. 55, n. 3, pp. 156–162, 2017.

[53] SUN, L., HOU, J., SHU, T. "Spatial and temporal contextual multi-armed bandit handovers in ultra-dense mmWave cellular networks", *IEEE Transactions on Mobile Computing*, v. 20, n. 12, pp. 3423–3438, 2020.

[54] 3GPP. *5G; NR; Radio Link Control (RLC) protocol specification* . Technical Specification (TS) 38.322, 3rd Generation Partnership Project (3GPP), 07 2020. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138300_138399/138322/16.01.00_60/ts_138322v160100p.pdf>. Version 16.1.0.

[55] 3GPP. *5G;NG-RAN;Architecture description.* Technical Specification (TS) 38.401, 3rd Generation Partnership Project (3GPP), 09 2018. Disponível em: <https://www.etsi.org/deliver/etsi_ts/138400_138499/138401/15.03.00_60/ts_138401v150300p.pdf>. Version 15.3.0.

[56] PARK, H.-S., LEE, Y., KIM, T.-J., et al. "Faster recovery from radio link failure during handover", *IEEE Communications Letters*, v. 24, n. 8, pp. 1835–1839, 2020.

[57] KODA, Y., NAKASHIMA, K., YAMAMOTO, K., et al. "Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning", *IEEE Transactions on Cognitive Communications and Networking*, v. 6, n. 2, pp. 802–816, 2019.

[58] MARTIKAINEN, H., VIERING, I., LOBINGER, A., et al. "On the basics of conditional handover for 5G mobility". In: *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–7. IEEE, 2018.

[59] DEB, S., RATHOD, M., BALAMURUGAN, R., et al. "Performance evaluation of conditional handover in 5G systems under fading scenario", *arXiv preprint arXiv:2403.04379*, 2024.

[60] CASTRO-HERNANDEZ, D., PARANJAPE, R. "Optimization of handover parameters for LTE/LTE-A in-building systems", *IEEE Transactions on Vehicular Technology*, v. 67, n. 6, pp. 5260–5273, 2017.

[61] FAROOQ, M. U. B., MANALASTAS, M., ZAIDI, S. M. A., et al. "Machine Learning Aided Holistic Handover Optimization for Emerging Networks", *arXiv preprint arXiv:2202.02851*, 2022.

[62] CHAIRMAN, T. "Summary of 3GPP TSG-RAN Workshop on Relseas 12 and Onward". In: *3GPP Workshop on Release*, v. 12, pp. 11–12, 2012.

[63] HSIEH, P.-J., LIN, W.-S., LIN, K.-H., et al. "Dual-connectivity prevenient handover scheme in control/user-plane split networks", *IEEE Transactions on Vehicular Technology*, v. 67, n. 4, pp. 3545–3560, 2017.

[64] ROSA, C., PEDERSEN, K., WANG, H., et al. "Dual connectivity for LTE small cell evolution: Functionality and performance aspects", *IEEE Communications Magazine*, v. 54, n. 6, pp. 137–143, 2016.

[65] OGUMA, Y., ARAI, R., NISHIO, T., et al. "Proactive base station selection based on human blockage prediction using RGB-D cameras for mmWave communications". In: *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. IEEE, 2015.

[66] WANG, X., KONG, L., WU, J., et al. "mmHandover: A Pre-Connection based Handover Protocol for 5G Millimeter Wave Vehicular Networks".

In: *2019 IEEE/ACM 27th International Symposium on Quality of Service (IWQoS)*, pp. 1–10. IEEE, 2019.

[67] IQBAL, S. B., KARABULUT, U., AWADA, A., et al. "Mobility Performance Analysis of RACH Optimization Based on Decision Tree Supervised Learning for Conditional Handover in 5G Beamformed Networks". In: *European Wireless 2023; 28th European Wireless Conference*, pp. 68–75. VDE, 2023.

[68] IQBAL, S. B., KARABULUT, U., AWADA, A., et al. "RACH-less Handover with Early Timing Advance Acquisition for Outage Reduction", *arXiv preprint arXiv:2403.10286*, 2024.

[69] ASHRAF, M. I., LIU, C.-F., BENNIS, M., et al. "Towards low-latency and ultra-reliable vehicle-to-vehicle communication". In: *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5. IEEE, 2017.

[70] VU, T. K., LIU, C.-F., BENNIS, M., et al. "Ultra-reliable and low latency communication in mmWave-enabled massive MIMO networks", *IEEE Communications Letters*, v. 21, n. 9, pp. 2041–2044, 2017.

[71] ZARIFNESHAT, M., ROY, P., XIAO, L. "Multi-Objective Approach for User Association to Improve Load Balancing and Blockage in Millimeter Wave Cellular Networks", *IEEE Transactions on Mobile Computing*, 2021.

[72] FENG, M., MAO, S., JIANG, T. "Dealing with Link Blockage in mmWave Networks: A Combination of D2D Relaying, Multi-beam Reflection, and Handover", *IEEE Transactions on Wireless Communications*, 2022.

[73] PRADO, A., VIJAYARAGHAVAN, H., KELLERER, W. "ECHO: Enhanced conditional handover boosted by trajectory prediction". In: *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 01–06. IEEE, 2021.

[74] ALRABEIAH, M., ALKHATEEB, A. "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels", *IEEE Transactions on Communications*, v. 68, n. 9, pp. 5504–5518, 2020.

[75] MARASINGHE, D., RAJATHEVA, N., LATVA-AHO, M. "LiDAR aided human blockage prediction for 6G". In: *2021 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6. IEEE, 2021.

[76] YAJNANARAYANA, V., RYDÉN, H., HÉVIZI, L. "5G handover using reinforcement learning". In: *2020 IEEE 3rd 5G World Forum (5GWF)*, pp. 349–354. IEEE, 2020.

[77] SUN, Y., FENG, G., QIN, S., et al. "The SMART handoff policy for millimeter wave heterogeneous cellular networks", *IEEE Transactions on Mobile Computing*, v. 17, n. 6, pp. 1456–1468, 2017.

[78] XU, C., LIU, S., ZHANG, C., et al. "Joint user scheduling and beam selection in mmWave networks based on multi-agent reinforcement learning". In: *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1–5. IEEE, 2020.

[79] ALKHATEEB, A., BELTAGY, I., ALEX, S. "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff". In: *2018 IEEE Global conference on signal and information processing (GlobalSIP)*, pp. 1055–1059. IEEE, 2018.

[80] ZHOHOV, R., PALAIOS, A., RYDÉN, H., et al. "Reducing Latency: Improving Handover Procedure Using Machine Learning". In: *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1–5. IEEE, 2021.

[81] DU, F., CHEN, G., QIU, L. "The Analysis of Mobility Performance in MmWave-μWave Heterogeneous Networks". In: *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 908–913. IEEE, 2019.

[82] AKDENIZ, M. R., LIU, Y., SAMIMI, M. K., et al. "Millimeter wave channel modeling and cellular capacity evaluation", *IEEE journal on selected areas in communications*, v. 32, n. 6, pp. 1164–1179, 2014.

[83] YOU, L., LIAO, Q., PAPPAS, N., et al. "Resource optimization with flexible numerology and frame structure for heterogeneous services", *IEEE Communications Letters*, v. 22, n. 12, pp. 2579–2582, 2018.

[84] ALSENWI, M., TRAN, N. H., BENNIS, M., et al. "eMBB-URLLC resource slicing: A risk-sensitive approach", *IEEE Communications Letters*, v. 23, n. 4, pp. 740–743, 2019.

[85] KARIMI, A., PEDERSEN, K. I., MOGENSEN, P. "5G URLLC performance analysis of dynamic-point selection multi-user resource allocation". In: *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 379–383. IEEE, 2019.

[86] CHARAN, G., ALRABEIAH, M., ALKHATEEB, A. "Vision-Aided 6G Wireless Communications: Blockage Prediction and Proactive Handoff", *IEEE Transactions on Vehicular Technology*, v. 70, n. 10, pp. 10193–10208, 2021. doi: 10.1109/TVT.2021.3104219.

[87] MURAKAMI, T., YOSHIKAWA, K., YAMAGUCHI, A., et al. "Blockage Prediction in an Outdoor mmWave Environment by Machine Learning Employing a Top View Image". In: *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2022. doi: 10.1109/PIMRC54779.2022.9977578.

[88] DEMIRHAN, U., ALKHATEEB, A. "Radar Aided Proactive Blockage Prediction in Real-World Millimeter Wave Systems". In: *ICC 2022 - IEEE International Conference on Communications*, pp. 4547–4552, 2022. doi: 10.1109/ICC45855.2022.9838438.

[89] NGO, H. Q., ASHIKHMIN, A., YANG, H., et al. "Cell-Free Massive MIMO Versus Small Cells", *IEEE Transactions on Wireless Communications*, v. 16, n. 3, pp. 1834–1850, 2017. doi: 10.1109/TWC.2017.2655515.

[90] NGO, H. Q., INTERDONATO, G., LARSSON, E. G., et al. "Ultra-dense Cell-Free Massive MIMO for 6G: Technical Overview and Open Questions", *Proceedings of the IEEE*, pp. 1–27, 2024. doi: 10.1109/JPROC.2024.3393514.

[91] BENEWAKE. "TF350 Ultra Long Range Single-point LiDAR". 2024. Disponível em: <`https://en.benewake.com/TF350/index.html`>.

[92] RIERA-PALOU, F., FEMENIAS, G. "Decentralization Issues in Cell-free Massive MIMO Networks with Zero-Forcing Precoding". In: *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 521–527, 2019. doi: 10.1109/ALLERTON.2019.8919893.

[93] PRANANTO, BAUD HARYO AND ISKANDAR AND KURNIAWAN, ADIT. "A New Method to Improve Frequent-Handover Problem in High-Mobility Communications Using RIC and Machine Learning", *IEEE Access*, v. 11, pp. 72281–72294, 2023. doi: {10.1109/ACCESS.2023.3294990}.

[94] LACAVA, A., POLESE, M., SIVARAJ, R., et al. "Programmable and Customized Intelligence for Traffic Steering in 5G Networks Using Open RAN Architectures", *IEEE Transactions on Mobile Computing*, v. 23, n. 4, pp. 2882–2897, 2024.

[95] RINALDI, F., RASCHELLA, A., PIZZI, S. “5G NR system design: a concise survey of key features and capabilities”, *Wireless Networks*, v. 27, n. 8, pp. 5173–5188, 2021.

[96] RANGAN, S., RAPPAPORT, T. S., ERKIP, E. “Millimeter-wave cellular wireless networks: Potentials and challenges”, *Proceedings of the IEEE*, v. 102, n. 3, pp. 366–385, 2014.

[97] YANG, B., YANG, X., GE, X., et al. “Coverage and handover analysis of ultra-dense millimeter-wave networks with control and user plane separation architecture”, *IEEE Access*, v. 6, pp. 54739–54750, 2018.

[98] GUROBI OPTIMIZATION, LLC. “Gurobi Optimizer Reference Manual”. 2022. Disponível em: <`https://www.gurobi.com`>.

[99] ALJERI, N., BOUKERCHE, A. “Mobility management in 5G-enabled vehicular networks: models, protocols, and classification”, *ACM Computing Surveys (CSUR)*, v. 53, n. 5, pp. 1–35, 2020.

[100] SAAD, W. K., SHAYEA, I., HAMZA, B. J., et al. “Handover parameters optimisation techniques in 5G networks”, *Sensors*, v. 21, n. 15, pp. 5202, 2021.

[101] O-RAN ALLIANCE. “O-RAN Alliance Home page”. 2024. Disponível em: <`https://www.o-ran.org/`>.

[102] WG1: USE CASES AND OVERALL ARCHITECTURE WORKGROUP. *Use Cases Analysis Report*. Technical Report (TR) R003, O-RAN Alliance, 06 2024. Disponível em: <`https://specifications.o-ran.org/download?id=638`>. Version 14.00.

[103] JIANG, C., ZHANG, H., REN, Y., et al. “Machine learning paradigms for next-generation wireless networks”, *IEEE Wireless Communications*, v. 24, n. 2, pp. 98–105, 2016.

[104] LI, T., SAHU, A. K., TALWALKAR, A., et al. “Federated learning: Challenges, methods, and future directions”, *IEEE signal processing magazine*, v. 37, n. 3, pp. 50–60, 2020.

Appendix A

Brute Force Validation and Model Complexity

To test and validate the optimization model, we developed a brute force algorithm. The result of the brute force algorithm must be the same as the optimization in terms of the objective function value. If the resulting objective function value of the brute force algorithm is greater than the optimization model and the output combination of handovers is valid, then the optimization model has been proved wrong. It is important to ensure that the optimization model and brute force algorithm are tested in the same conditions, for instance, the objective value must be calculated in the same way. The Algorithm (4) describes the brute force algorithm main part.

Algorithm 4: Brute force validation algorithm

```
perm_length ← ⌊T/τ⌋
permutations ← all permutations of BS with length up to perm_length
for each permutation in permutations do
    obj_values ← BspermutationAssessment(permutation, 0, T)
end
best_permutation ← argmax_{c ∈ permutations} {obj_values(c)}
```

The brute force algorithm generates all the possible permutations of $|M|$ BS with length up to $\lfloor T/\tau \rfloor$ without consecutive repetitions, where $\lfloor . \rfloor$ is the rounding floor function. In the worst case, a TTT would be triggered in the first timeslot where the UE associates with the new BS, resulting in a new handover at every τ timeslots. Then, the maximum length is equal to $\lfloor . \rfloor$ of the quotient between T, the observation time, and τ, the TTT.

Lemma A.0.1. *The number of BS permutations, without consecutive repetition and*

with length not greater than $\lfloor T/\tau \rfloor$ *is*

$$|M| \sum_{i=1}^{\lfloor T/\tau \rfloor} (|M|-1)^{i-1} \tag{A.1}$$

Proof. Consider E elements grouped in sequences of length k with repetition, except for consecutive elements. This means that an element can be repeated in the sequence, only if it is not equal to its predecessor, for instance, the $i-th$ element of the sequence must be different from the $(i-1)-th$ element. At the first position of the sequence, there are E elements to pick, but in the $k-1$ consecutive ones there will always be $E-1$. So, the permutations $P(k)$

$$\begin{aligned}
&k = 1 \rightarrow P(1) = E \\
&k = 2 \rightarrow P(2) = E(E-1) \\
&k = 3 \rightarrow P(3) = E(E-1)(E-1) = E(E-1)^2 \\
&\quad \vdots \\
&k = n \rightarrow P(n) = E(E-1)(E-1)\dots(E-1) = E(E-1)^{n-1}
\end{aligned}$$

Summing all $P(k)$ such that $k \leq n$, we have the total number of sequences S

$$S = \sum_{k=1}^{n} E(E-1)^{k-1} = E \sum_{k=1}^{n} (E-1)^{k-1} \tag{A.2}$$

□

Subsequently, each permutation is individually evaluated considering the objective function value it would result in. This evaluation is done by the routine `BspermutationAssessment(.)`, described in Algorithm (5). This recursive routine receives as an argument the BS permutation to be evaluated, and the starting and ending timeslots of the evaluation.

The recursive routine `BspermutationAssessment(.)`, for every permutation with a length greater than 1, takes the first and the second element of the permutation as the SBS and TBS, respectively. Then, every timeslot is checked for a handover, and if that happens the recursion is called. Each recursion level will return the maximum objective function value between the two BS for the time interval $[t_i, t_f]$. Considering that BS A offers the highest objective function value at t_i, if there are handover opportunities to BS B from t_i+1 to t_i+k, then each of these opportunities will be individually evaluated by the loop in line 6 and will only be accepted as best value if the condition on line 11 is true. On the other hand, assuming it is not worth to handover to BS B, it will not meet the condition in line 11 for $t < t_i + 1$, or a

Algorithm 5: BS permutation assessment

```
1  Function BspermutationAssessment(permutation, init, end)
2      l = length(permutation)
3      if l > 1 then
4          serving_bs ← permutation[0]
5          target_bs ← permutation[1]
6          for t in [t_i, t_f] do
7              if Eq. (3.3) is true for p =serving_ bs and q =target_ bs then
8                  current_value ← Σ_{i∈[t_i,t]} S^p_{n,i} x^p_{n,i} + I_{n,i} y^p_{n,i}
9                  current_value ← current_value +
                    BspermutationAssessment(permutation[q:], t, T)
10             end
11             if current_ value > obj_ value then
12                 obj_value ← current_value
13             end
14         end
15     end
16     return obj_ value
17 end
```

higher value for $t > t_i + k$ will satisfy the condition further. Then, the recursive routine `BspermutationAssessment(.)` will always return the best possible objective function value for a given permutation. All the permutations are evaluated in the line 6 of Algorithm (4), guaranteeing that the best objective value will be found between the local maxima associated to each permutation.

Lemma A.0.2. *Each call to* `BspermutationAssessment(.)`*, passing a permutation of length l, has the number of iterations equal to*

$$\prod_{j=1}^{l-1}(T - t_i - j\tau) \quad \forall\ l < \left\lfloor \frac{T - t_i}{\tau} \right\rfloor,\ 0 < t_i < T \tag{A.3}$$

Proof. In the interval $[t_i, T]$, the earliest handover opportunity is at $t_i + \tau$. Hence, $T - t_i - \tau$ timeslots need to be evaluated. If a second handover might happen, then it may be triggered from $t_i + 2\tau$ on, resulting in $T - t_i - 2\tau$ timeslots to be evaluated, and so successively. Considering the two handovers, the permutation of timeslots for evaluation is $(T - t_i - \tau) \times (T - t_i - 2\tau)$. This is true for every sequence of BS association of length l, since $l < \lfloor (T - t_i)/\tau \rfloor$. Otherwise, the product on Eq. A.3 would violate the maximum number of handovers on an interval, that is limited by τ, and results in a negative number of iterations, which is a contradiction. □

This algorithm is impossible to solve by enumeration in a feasible time for large values of T. Indeed, the worst-case complexity is given by the product between

Equations (A.1) and (A.3). The resultant complexity is expressed in Equation (A.4). Finally, we can assume that brute force is an exhaustive search on the solution space, once every possible combination is realized and they are all compared together at the end of the Algorithm (4), indicating the brute force optimality.

$$\mathcal{O}\left(|M|\sum_{i=1}^{\lfloor T/\tau \rfloor}\left[(|M|-1)^{i-1}\prod_{j=1}^{i-1}(T-j\tau)\right]\right) \tag{A.4}$$