Master's Thesis

Optimizing Allocation and Handover Processes in Mobile Networks

Geymerson dos Santos Ramos
geymerson@ic.ufal.br

Advisors:

Dr. André Luiz Lins de Aquino
Dr. Rian Gabriel Santos Pinheiro

Maceió
October 27, 2021

Geymerson dos Santos Ramos

Optimizing Allocation and Handover Processes in Mobile Networks

A thesis submitted by Geymerson dos Santos Ramos in partial fulfillment of the requirements for the degree of Master of Science in Informatics at the Federal University of Alagoas, Computing Institute.

Advisers:

Dr. André Luiz Lins de Aquino

Dr. Rian Gabriel Santos Pinheiro

Maceió
October 27, 2021

A thesis submitted by Geymerson dos Santos Ramos in partial fulfillment of the requirements for the degree of Master of Science in Informatics at the Federal University of Alagoas, Computing Institute, approved by the examination committee which signs bellow.

Dr. André Luiz Lins de Aquino - Adviser
Computing Institute
Federal University of Alagoas

Dr. Rian Gabriel Santos Pinheiro - Co-Adviser
Computing Institute
Federal University of Alagoas

Dr. Alexandre Mendes - Committee Member
School of Information and Physical Sciences
University of Newcastle

Dr. Erick de Andrade Barboza - Committee Member
Computing Institute
Federal University of Alagoas

Dr. Marilia Pascoal Curado - Committee Member
Department of Informatics Engineering
University of Coimbra

Maceió
October 27, 2021

Catalogação na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

R175o Ramos, Geymerson dos Santos.
Optimizing allocation and hadover processes in mobile networks / Geymerson dos Santos Ramos. – 2021.
49 f. : il.

Orientador: André Luiz Lins de Aquino e Rian Gabriel Santos Pinheiro.
Dissertação (mestrado em Informática) - Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2021.

Bibliografia: f. 44-49.

1. Redes móveis. 2. Alocação de usuários. 3. Otimização de processos. 4. Handover. 5. Redes locais sem fio. I. Título.

CDU: 004.72

UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Informática – PPGI
Instituto de Computação/UFAL
Campus A. C. Simões BR 104-Norte Km 14 BL 12 Tabuleiro do Martins
Maceió/AL - Brasil CEP: 57.072-970 | Telefone: (082) 3214-1401

Folha de Aprovação

GEYMERSON DOS SANTOS RAMOS

OTIMIZAÇÃO DE PROCESSOS DE ALOCAÇÃO E HANDOVER EM REDES MÓVEIS

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas e aprovada em 27 de outubro de 2021.

Banca Examinadora:

Prof. Dr. ANDRE LUIZ LINS DE AQUINO
UFAL – Instituto de Computação
Orientador

Prof. Dr. ERICK DE ANDRADE BARBOZA
UFAL – Instituto de Computação
Examinador Interno

Prof. Dr. RIAN GABRIEL SANTOS PINHEIRO
UFAL – Instituto de Computação
Coorientador

Prof. Dr. ALEXANDRE MENDES
Examinador Externo

Prof. Dra. MARILIA PASCOAL CURADO
Examinador Externo

Acknowledgement

Kill the boy, Jon Snow [...]. Kill the boy and let the man be born. – Aemon Targaryen to Jon Snow (George R.R. Martin, A Dance with Dragons).

For the last months, I have been fighting against emotions that I can only describe as anxiety due to fear of change. This work should have been done a long time ago. The plan for the past two years was simple: Before leaving home, I wanted to create my financial protection for study expenses, help my parents with their house renovation, present my thesis, and start a Ph.D. But then, the pandemic hit. I am feeling very compelled not to write about the covid-19, but it has inflicted massive pain and distress for people. I can not ignore the context in which this work is being conceived. I am a very fortunate person, for all my family made it through and is doing well. My friends are healthy, and I am also healthy... physically. The pandemic brought us dread and isolation. Doubt and uncertainty slowly infiltered me over time.

My plans seemed to make no sense anymore, something from a distant reality. The world was closed, and I was worried about the growing disconnection with my peers, my eagerness for learning, and the feeling of overcoming obstacles. Helplessness everywhere and idleness through the day. These were the gates to the transition to the void, and I needed to get out of there. Fortunately, I achieved most of my initial goals and took the time to make amends with myself. Things seem to be in motion again, and I still have some work in progress.

It is 1:56 am, and I need to end this monologue and go to sleep. But not before expressing my gratitude to Professor Rian Pinheiro, which I only met at the end of my undergraduate course but proved to be very important through the following years. And once again, thank you too, Professor André Aquino, the man who will not let me lose sight of my goals. Gratitude to my parents Cícera and Sandro, who raised and supported me to be who I am. There is no point in postponing it further. I recognize that there is apprehension, almost fear. Fear of the path that I need to follow. Fear of cutting the cord. It seems like it is time to go. May the light guide me through my journey, may my will be reforged from steel to gold. Let the change embrace me. Kill the boy.

Digna Factis Recipimus.

– Luke 23:41

Resumo

O número crescente de dispositivos conectados à Internet tem exigido avanços em tecnologias de comunicação sem fio. A rede 4G e suas antecessoras estão sendo gradativamente substituídas pela 5G, que promete maior velocidade, heterogeneidade e escalabilidade. A 5ª geração oferece suporte amplo para aplicações de redes definidas por software, aumentando a flexibilidade para modelagem de processos e protocolos que antes eram embarcados e de difícil atualização.

Este trabalho tem como objetivo melhorar processos em redes móveis através de modelos matemáticos, que podem impactar mobilidade, balanceamento de carga na rede e redução de custos operacionais. Nossa proposta visa a alocação de usuários em torres ou estações bases de redes de telecomunicação, minimizando *handovers* e melhorando a qualidade de comunicação.

O trabalho oferece as seguintes contribuições: i) Um modelo matemático para alocação de usuários em estações bases de redes de telefonia móvel, com a redução de transferências; ii) Uma solução meta-heurística como alternativa a modelos exatos, visto que estes podem se tornar inviáveis em condições de restrição de recursos de tempo e computacionais; iii) A avaliação dos modelos em cenários simulados de mobilidade, avaliando o processo de *handover* e a distribuição de usuários na rede em função de largura de banda disponível.

A modelagem, que considera a frequência média de *handover* de cada estação base e o sinal indicador de qualidade de comunicação, foi avaliada com soluções exatas e heurísticas, sendo estas o algoritmo de *branch and bound*, busca local iterativa, e solução gulosa. Através dos métodos heurísticos o algoritmo de busca local iterativa obteve uma redução de aproximadamente 82% do tempo de execução em comparação ao algoritmo exato *branch and bound*. Com relação ao indicador de qualidade de conexão, a solução obteve um ganho médio de 1.45% em comparação à solução da literatura, mantendo o número *handovers*. Apesar do ganho reduzido, o que torna nossa proposta estatisticamente equivalente, oferecemos a vantagem de não computar todas possíveis e futuras rotas dos usuários, sendo suficiente a posição atual. Adicionalmente, nossa solução considera a capacidade de largura de banda de cada estação base, respeitando a capacidade de rede e mantendo o controle de alocação.

Palavras-chave: Redes Móveis, Alocação, Otimização, *Handover*.

Abstract

The growing number of devices connected to the Internet has required advances in wireless communication technologies. As a result, 5G networks gradually replace 4G and its predecessors, offering more speed, heterogeneity, and scalability. The fifth-generation provides broad support for software-defined networking (SDN) applications, increasing the programming flexibility of processes and protocols previously embedded and difficult to update.

This work aims to improve processes in mobile networks through mathematical models. Our work focuses on optimizing the allocation of users in base stations of telecommunication networks, minimizing the handover of users between base stations, and improving network communication quality.

The contributions of this work are: i) A mathematical model for allocating users of mobile networks at base stations, also aiming handover reduction; ii) A metaheuristic solution as an alternative to exact models since exact models can prove to be non-scalable and present unfeasible solving times under computationally restricted conditions; iii) A model evaluation in simulated mobility scenarios considering the handover process and the network user distribution according to available bandwidth.

Our allocation model considers the average handover frequency of each base station and the Reference Signal Received Quality (RSRQ) indicator between users and base stations. The model evaluation used exact and heuristic methods: the branch and bound algorithm, iterated local search, and a greedy solution. On average, the iterated local search algorithm obtained an execution time reduction of approximately 82% compared to the branch and bound exact algorithm. Regarding the RSRQ indicator, the solution reached a 1.45% average gain, and the number of performed handovers was maintained, compared to a similar literature model. Despite the modest improvement, which makes our proposal statistically equivalent to the literature model, we offer the advantage of not predicting the users' possible and future routes. Only the current position is required. Furthermore, our solution also considers base stations' bandwidth capacity, controlling the allocation and network occupation limits.

Keywords: Mobile Networks, Allocation, Optimization, Handover.

Contents

Abbreviation List

CAGR	Compound Annual Growth Rate
D2D	Device-to-Device
DC	Data Center
eNB	evolved Node B
EPS	Evolved Packet System
GAP	Generalized Assignment Problem
GPS	Global Positioning System
GRASP	Greedy Randomized Adaptive Search Procedure
HSS	Home Subscriber Server
ILP	Integer Linear Programming
ILS	Iterated Local Search
IoT	Internet of Things
LTE	Long Term Evolution
LP	Linear Programming
LPP	Linear Programming Problem
LPWA	Low-Power Wide-Area
M2M	Machine-to-Machine
MBS	Macro Base Station
MME	Mobile Management Entity
PDN-GW	Packet Data Network Gateway
RAT	Radio Access Technology
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indication
SBS	Small Base Station
SDN	Software-Defined Networking
S-GW	Serving Gateway

SUMO	Simulation of Urban MObility
UE	User Equipment
VND	Variable Neighborhood Descent

List of Figures

List of Tables

List of Algorithms

1

INTRODUCTION

The evolution of communication technologies has significantly impacted protocols and data transfer solutions. With the arrival of the Internet of Things (IoT) (Al-Fuqaha et al., 2015) paradigm, devices are becoming more integrated, and Wireless Networks (Akyildiz et al., 2002) are showing to be essential resources. Through devices such as smartphones, users can communicate from practically anywhere, make use of location services, and exchange information with technologies such as Wi-Fi, Bluetooth, 3G, 4G, and the most recent 5G (Panwar et al., 2016). The fifth-generation promises more incredible speeds, ubiquitous connection, support to denser, heterogeneous, and scalable networks. Among the fundamentals pillars of 5G networks, we highlight:

- **Uninterrupted and ubiquitous connection:** Users will able to connect from anywhere at any time;
- **Zero-latency:** Granted support to real-time applications, critical systems, and services with low tolerance to network delay;
- **High-speed connections:** Support to applications that rely on virtually zero latency. The connection speed will be in the order of Gigabit per second.

The new scenario offers the opportunity for innovation, improvement of existing solutions, and the use of new computational paradigms, such as Software-Defined Networking (SDN) (Nunes et al., 2014), which allows flexible programming and management of resources such as scalability, privacy, security, traffic control, and allocation. In the current state of mobile wireless networks, the 5G technology will demand applications to be faster, more efficient, heterogeneous, and scalable.

Existing solutions will require improvement or rethinking to a zero-latency context. The handover process is one of the many problems that we can approach and propose several improvements. Figure 1.1 shows a handover process (Yan et al., 2010). It occurs when a user connected to some device (UE — User Equipment) changes its base station (here represented by the evolved Node B — eNB). This process can affect mobility, connection discontinuity perception, and commutation between different wireless communication technologies such as Wi-Fi, 4G, or Bluetooth. Therefore, handovers need to occur imperceptibly, within short time intervals, and in areas that do not negatively affect mobility and connectivity experience.

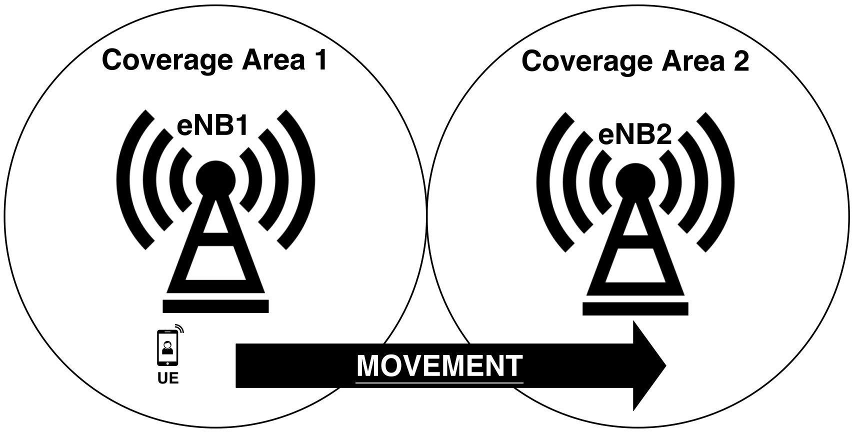


Figure 1.1: Handover of UE from eNB_1 within coverage area 1 to eNB_2 in the coverage area 2.

Our work is motivated by the current state of mobile networks, which is transitioning to a technology (5G) with a growing adoption of optimization models run through SDN applications, which consider a logically centralized controller to execute models that dynamically optimize, manage, and improve network features such as mobility and load distribution.

Using mathematical models, we aim to optimize the handover and allocation of users in base stations of telecommunication networks. As a starting point, our proposal extends a model that minimizes both handovers between base stations and the communication cost between base stations and data centers (Taleb et al., 2015). The authors did not consider the direct relation user-base station, so we extended their model by proposing an optimization solution to allocate users. We later compare our proposal to a similar literature model (Ahmadi et al., 2020). We do not evaluate our solution in wireless SDN controllers, but it is appliable in this context.

The resulting allocation model first considered the average handover frequency of each base station and their distance from users. However, since we deal with wireless communication, the distance parameter can be non-responsive to noise and environment interference. We updated the allocation model by replacing the distance parameter with the RSRQ (Reference Signal Received Quality), which indicates the quality of communication between users and base stations.

We evaluate the allocation models in two simulated mobility scenarios. The first considers a 13.7 km route at Maceió City, Brazil, with eNBs artificially placed. The distance-based model computes the distance and allocates the user in base stations according to the latitude and longitude of GPS (Global Positioning System) collected data. In the second scenario, we use a heuristic approach to solve the model instead of an exact algorithm, and modify the model to consider the RSRQ indicator instead of distance. We use a map region of the city of São Paulo, Brazil. The map region has the real position of 26 eNBs from a local phone carrier, provided by

Telebrasil (2021), and we simulate the user routes with the SUMO (Simulation of Urban MObility) simulator (Krajzewicz et al., 2012).

In our final evaluations, on average, the iterated local search algorithm obtained an execution time reduction of approximately 82% compared to the branch and bound exact algorithm. Comparing results with Ahmadi et al. (2020) in the second scenario, the RSRQ indicator reached a 1.45% average gain, and the models performed the same number of handovers.

We found in the literature several related works about the topic of this work. For instance, Kuklinski et al. (2015) present a discussion on how to apply technologies such as SDN to increase the efficiency of mobile networks. They consider the handover process according to 3 architectures: i) *centralized*: uses one controller; ii) *semi-centralized*: multiple controllers acting on different domains or geographic regions; iii) *hierarchical*: multiple layers and a master controller on the top of the hierarchy that communicates with other controllers on the lower levels.

Prados-Garzon et al. (2016) propose a handover implementation for a partially virtualized LTE (Long Term Evolution) networks. The authors implemented the process's message exchange, simulated the transmission delays, propagation, network processing, and handover finalization.

Lee and Yoo (2017) present a 5G handover scheme considering users' mobility information. The controller receives mobility data and base station information to select in which cell to allocate the users. They observed that the proposal could select the cells that offer greater signal strength and fit the user's movement direction through simulation.

Qiang et al. (2016) discuss a solution to a handover multi-objective problem for hybrid 5G environments. The authors consider maximizing the data receiving rate and minimizing the probability of handover process failure. The probability is calculated based on other users of the network and limited information (due to privacy restrictions). In a previous work (Qiang et al., 2014), a controller executes this process using private information.

Duan and Wang (2015) proposal focuses on privacy and authentication tools for 5G SDN heterogeneous networks handover processes. Authentication uses attributes such as location, direction, and features of the physical network layer to generate unique identifiers. Cryptography is not required, thus simplifying the authentication process.

This work will mainly contribute with the following aspects:

- Proposition of a mathematical allocation model to allocate users in base stations of mobile networks;
- Proposition and evaluation of heuristic solving methods to the allocation model as an alternative to exact solving approaches;
- Evaluation of allocation on different mobility simulations, one of which considers the real position of base stations from a local mobile carrier;
- A evaluation and comparison of our model with existing literature work.

Additionally, some results of our proposal have already been published and are available (Ramos et al., 2019b,a).

The remainder of this work is structured as follows: Chapter 2 presents general concepts on mobile networks, handover and optimization. Chapter 3 shows the proposal, which contains description of our allocation models, mainly the distance-based model, the RSRQ allocation model, and the heuristic solution. Next, Chapter 4 discusses the results. Finally, Chapter 5 presents the final considerations and future work.

This chapter briefly introduced relevant topics such as mobile networks and what is the handover process. We also discussed objectives, related literature, and contributions of this work.

2

TECHNOLOGIES AND CONCEPTS

This chapter presents the concepts and technologies related to our work. In Section 2.1, we will discuss wireless technologies such as 5G and its predecessors. Section 2.2 presents an overview of handover processes. We end the chapter with Section 2.3 discussing optimization, mainly the algorithms and methods which contribute to the understanding of our proposal.

2.1 Mobile networks

The number of connected devices is increasing, and due to the aggressive volume of data, we also observed the surging of new computational paradigms, such as Cloud Computing (Mell and Grance, 2010) and Big Data (Chen et al., 2014), to redesign data storage and analysis. Wireless communication technologies, which also should support applications in this new panorama, have found themselves in a hostile scenario against dense, heterogeneous, and high connectivity demands.

In such a scenario, the 5th-generation introduces itself to mobile communication. The 5G networks will allow the deployment of computational paradigms that will help the expansion of information and communication technologies, which are very important for Smart Cities (Mohanty et al., 2016). If compared to the 4th-generation, we can list the following improvements: 100 times more connected devices; 1000 times more data volume from connected devices; 100 times faster; 1-millisecond latency; almost 100% coverage; information processing in real-time and reduced energy consumption (GSMA Intelligence, 2014). In addition, the fifth generation is coming to significantly reduce operational infrastructure costs by offering ubiquitous connection, scalability, heterogeneity, and software solutions.

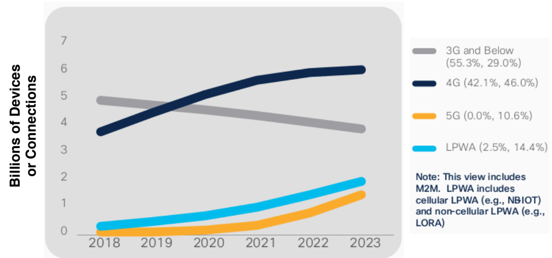
Table 2.1 shows a comparison of the five generations of mobile technologies. 1G mostly offered voice communication support through analog signals. Handover solutions were precarious and horizontal, which means transfers between access points did not happen between different communication technologies (e.g., 1G and 2G). Besides slow speed, the first generation had severe security problems because of the reproduction of conversations in the base stations, which made them susceptible to external access (Vora, 2015). 2G technology surpassed the previous generation's speed, adopted digital signals, images, and text messages. Its obsolescence became apparent as the Internet and the dissemination of multimedia data gained popularity, starting the 3G era.

Table 2.1: Comparative of the different generations of wireless communication technology.

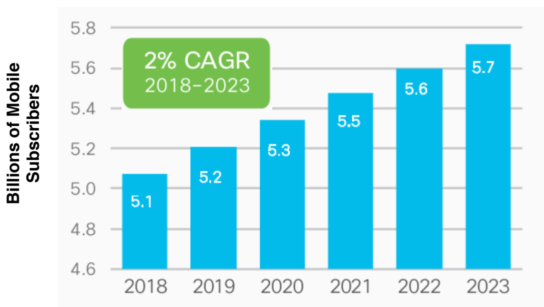
Generation	Features	Speed	Latency (ms)	Handover	Limitations
1G	Analog signals, voice message	< 2.4 Kbps	-	Horizontal	Low security
2G	Digital signal, voice messaging, text e images	< 64 Kbps	< 1000	Horizontal	weak support to Internet service
3G	Voice message, Access to home and mobile Internet, Video calls	< 3.1 Mbps	< 500	Horizontal and Vertical	Slow speed
4G	Greater speed connection	< 300 Mbps	< 100	Horizontal and Vertical	Low scalability and connectivity, support to dense networks
5G	Ubiquitous connection and high speed, scalability, heterogeneous	> 1 Gbps	< 1	Horizontal and Vertical	-

In the 3G era, the connections' speed reached the megabit per second order, and the Internet gained access through cell phones. As a result, multimedia streaming services increased significantly. Besides the horizontal approach, we found a new vertical architecture regarding handover processes, which allowed transfers between different wireless communication technologies (e. g. 3G to 2G). The following demands incurred on the network's speed, which resulted in the development of the 4G. Since then, the fourth generation became the main access option, reaching theoretical speeds of 300 Mpbs.

According to CISCO's annual internet report (CISCO, 2020) shown in Figure 2.1(a), the number of 4G network users will surpass the previous generation after 2019. The projection estimates that the fourth generation will correspond to 46.0% of connections by the year 2023. The figure also includes LPWA connections (Low-Power Wide-Area), essential in M2M communication (Machine-to-Machine).



(a) Global mobile device and connection growth.



(b) Global mobile subscriber growth.

Figure 2.1: Global projection for the number of devices and connections according to network type and for the global number of mobile subscribers. Adapted from (CISCO, 2020).

Figure 2.1(b) shows the growth and projection of mobile subscribers between 2018 and 2023 with a 2% CAGR (Compound Annual Growth Rate). These increasing numbers put pressure on the infrastructure of 4G networks, already being replaced by the upcoming technology, which is 5G.

Compared with its predecessor, the fifth generation is not simply a speed increase. It is a redefinition in applications, efficient use of resources, models, and architectures to support data loads and various devices with different radio communication technologies. It will support heterogeneous frameworks with software-defined networks, D2D (Device-to-Device), and full-duplex radio communication (Panwar et al., 2016). Figure 2.2 illustrates an example of the plurality that 5G network architectures can achieve. This figure also represents a multi-layered architecture.

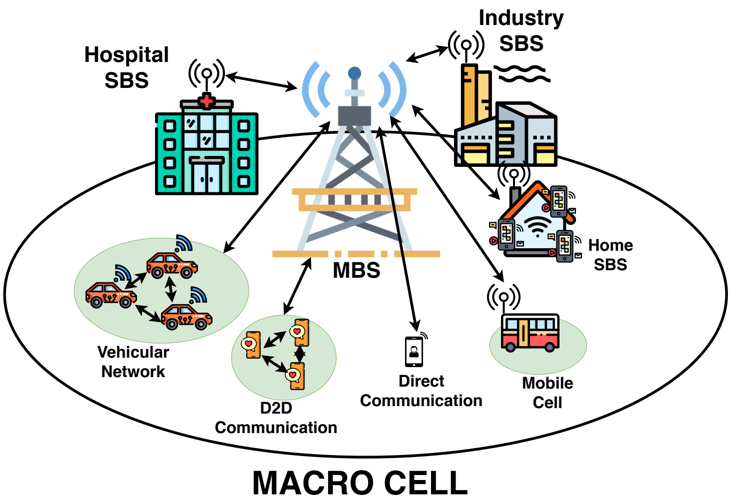


Figure 2.2: 5G network architecture composed of a macro cell, which contains stationary and mobile small cells, direct and D2D communication.

In this architecture, a macro cell with an MBS (Macro Base Station) in the upper layer receives requests from the lower layers, divided into small cells containing SBS (Small Base Stations) for different application profiles. For example, in a small mobile cell, its occupants can access the external network through an SBS attached to the bus, which communicates with the MBS of the macrocell. In addition, mobile devices such as smartphones can communicate directly with the MBS or create small dynamic D2D communication cells, in which only some of these devices connect to the MBS while the others transmit requests and responses.

SBS can be installed in diverse environments such as homes, hospitals, and industries, constituting different natures of applications and small cells. Depending on the radius range and number of users, cells can be distinguished as presented in Table 2.2.

The use of small cells increases macro cells range, and it also:

- **Increases transfer rates**: A local SBS, which communicates with an MBS, serves better the indoor devices reducing the signal attenuation effects that would occur in direct communication between MBS and devices;

Table 2.2: Range and total users capacity according to cell types. Adapted from Panwar et al. (2016)

	Femtocell	Picocell	Microcell	Macrocell
Range	10 - 20 m	200 m	2 km	30 - 35 km
Users	< 20	20 - 40	> 100	Many

- **Improves the use of the radio communication spectrum:** There are fewer devices in direct communication with MBS;
- **Increases energy efficiency**: Communication intermediated by an SBS reduces the required range for data transmission, which significantly impacts the battery usage of the devices;
- **Reduces cost**: An SBS has lower installation and operating costs if compared to an MBS.

We can mention the optimized use of computational and energy resources among the problems to be addressed by the fifth generation. Base stations operate at peak periods in the existing mobile phone networks, and their processing power is only for connected users. Energy consumption for peak or near-inactivity times is similar (Correia et al., 2010), increasing the network's operating costs.

Base stations in commercial areas get overloaded during the daytime, while residential areas have little activity (Wu et al., 2015). The scenario reverses for different day times. Instead of being restricted to connected users, we can achieve better usage and load-balancing by distributing idle computing resources.

Also, there are solutions for the redundant use of two channels (operating at different frequencies) for communicating with base stations. The devices have a channel transferring data (uplink) to the base station and another transfer channel in the opposite direction (downlink). This practice is considered inefficient. 5G networks present full-duplex communication using a single channel for sending and receiving data, with no co-interference.

Until now, mobile network architectures have not distinguished between indoor and outdoor users. As a result, devices communicate directly with the base stations, regardless of location, which implies severe attenuation and loss of signal quality if walls or obstacles surround the user. The fifth-generation introduces the SBS as a bridge between users and MBS, allowing indoor and outdoor distinction and improving transmission quality.

As for heterogeneous wireless networks (Damnjanovic et al., 2011), Bluetooth, 3G, and Wi-Fi, 5G's solution framework proposes significant contributions. 4G technology already provides some support, but not as a primary purpose since a device can only connect the uplink and downlink channels to the same base station.

The fifth-generation must implement decoupling, allowing users to have the channels associated with different base stations. According to Boccardi et al. (2016), decoupling reduces required transmission power, interference, achieves higher transmission rates, reduces costs, and differentiates load balancing for uplink and downlink channels.

Mobility and handover processes also require attention because transfers between different wireless technologies have become common nowadays. When moving between commercial and residential areas, users' connection must be sustained, with no perception of discontinuity or drop in transfer rates. 5G solutions must guarantee user services, even in high mobility scenarios with speeds of up to 500 km/h (Zhang et al., 2017).

It is noticeable that mobile wireless networks are converging for a highly connected ecosystem, in which users seamlessly transit between environments and communication technologies. We aim to contribute to this integrated ecosystem by improving the allocation and connection transitioning processes, namely handovers. In our experiments, we will consider simulated scenarios of mobility in which users are transferred regularly between base stations. As long as the technology can provide the input data, our model will calculate which base station offers the best service. We do not distinguish wireless cell types (small or macro) or base stations, but our solution ensures that each eNB has enough bandwidth for the requiring users.

2.2 Handover

The handover process is the transferring of a user connected to one access point to another. In general, we divided the process into 3 phases: preparation, execution, and completion. The preparation includes the analysis of a series of user information, such as position, services consumed, possible future routes, neighboring access points, and provided services' quality.

The information allows the best access points to be selected. Different solutions apply machine learning techniques as auxiliary selecting tools (Al et al., 2016). In the execution phase, the solutions use messages for synchronization and recognition elements involved in the handover process: users, access points, and upper-level entities that generally coordinate the process.

The types of handover can also vary. In 5G networks, it is possible to execute transfer between two SBS belonging to the same macro cell, between different macrocells, and even to another type of radio communication technology, as shown in Figure 2.3. If the base stations are in the same macro cell, the handover is an intra-macro cell. This process is the case of a transfer between SBS1 and SBS2. When transferring from SBS2 to SBS3, the device migrates to another cell and performs an inter-macro cell handover. When the device disconnects from the 5G network towers to another type of communication technology, such as 4G, we have the multi-RAT (Radio Access Technology) handover.

Intra-macro cell transfers have a lower cost when compared to inter-macro cells because only SBSs are involved in the process. However, at the inter-macro cell level, not only do SBSs need to be analyzed, but MBSs are also involved, and this consequently increases the complexity of handover, which can become even more costly for multi-RAT transfers.

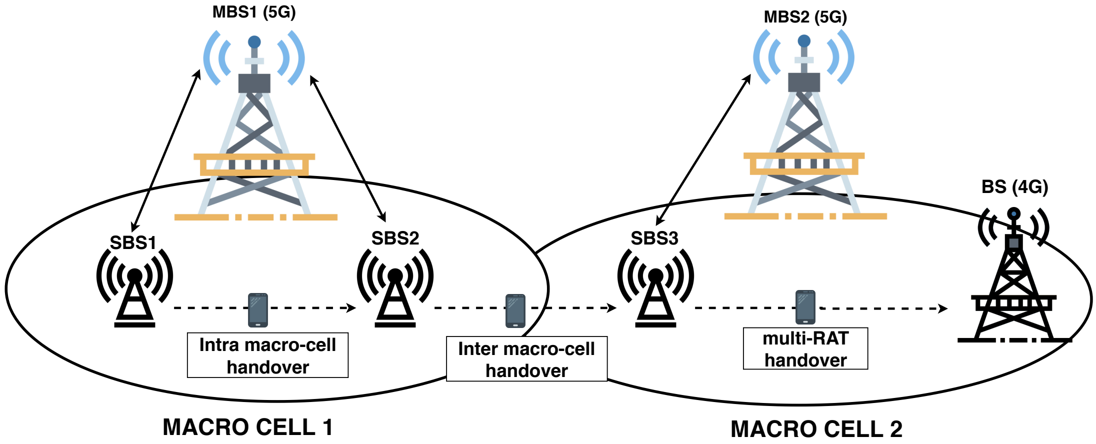


Figure 2.3: Intra macro-cell, inter macro-cell and multi-RAT handover.

Considering that we have the required data, such as user and base station location, the list of allowed eNBs, RSRQ values, etc., our model determines the most suited or the best allocation. There is no need to know eNBs technology types (multi-RAT handover) or handover level (intra-cell or inter-cell). Knowing and handling different types of wireless technologies, the management, preparation, and execution phases is a task of the network entity (such as an SDN controller) which executes our allocation model.

2.3 Optimization

The urge to increase efficiency, or more specifically, optimize a process, seems only natural after the discovery or creation stage. Our cognitive capacity and math domain has led us to maximize gains and minimize costs, such that optimization became a whole study field. When proposing an efficient solution to a problem, one must consider the most relevant variables, correlations, and constraints.

Different classes of problems require distinct computational effort to solve. Figure 2.4 shows a Venn diagram of the classes and their solving difficulty in regards to $\mathcal{P}$. If we consider the input size n of a given problem $p \in \mathcal{P}$ and some constant k, then p is solvable with a polynomial-time complexity $O(n^k)$ (Cormen et al., 2009, chapter 34). There is no known efficient solution for the problems in the $\mathcal{NP}$-hard class, and $\mathcal{P} \neq \mathcal{NP}$ or $\mathcal{P} = \mathcal{NP}$ is an assumption yet to be proven. Garey and Johnson (1979) provides a more detailed discussion of the problem classes and NP-Completeness.

We can achieve optimization through exact or heuristic methods. Exact methods are guaranteed to find an optimal solution, while heuristic methods are more flexible, fine-tuned, and specific to a given problem. It can find solutions that may not be global optima, but good enough if we consider the available resources. Metaheuristics, similarly to heuristics, do not guarantee a globally optimal solution but are more general solving problem frameworks.

The following topics of this section discuss relevant algorithms to this work, such as the simplex method, the generalized assignment algorithm, branch and bound, and iterated local search.

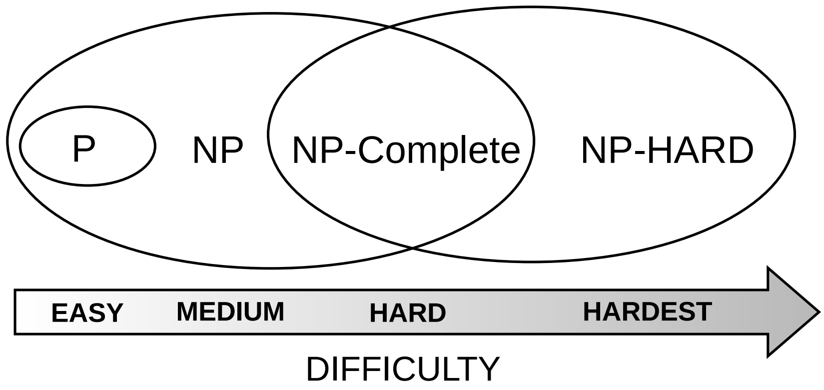


Figure 2.4: The relation between problem classes and complexity.

2.3.1 Linear programming and simplex method

The Simplex Algorithm solves Linear Programming Problems (LPP), widely found in real-world industry and business models. Linear optimization models play a major role in minimizing processes' expenses and maximizing gain. The general structure of an LPP contains three important components:

1. **Decision Variables:** It is a set of controllable, continuous, and non-negative values. Such variables guide the decisions and courses of action of an optimization model. For example, the employees of a factory and their respective tasks can be considered decision variables in an optimization model that aims to maximize the number of tasks per employee.

2. **Objective function:** It is a mathematical function containing a set of constants and the decision variables. Depending on the problem, the optimal value of this function can represent the minimum cost, maximum profit, and best efficiency.

3. **Constraints:** It is a set of mathematical expressions that limits the number of valid solutions. For example, a business analyst may create an objective function that models profit according to available resources and produced products. As a constraint, the available budget only allows the purchase of a limited number of resources. This process limits the number of produced products and maximum profit.

Equation (2.1) (Sharma, 2017, chapter 2) defines the relationship between these three main components. Z is the objective function that we aim to maximize or minimize, c_j is the cost or profit value associated with the decision variable x_j. The input-output coefficient (or technological coefficients) of the constraints' inequations are represented by a_{ij}, and b_i establishes lower or upper bounds for each constraint inequation. All the decision variables must be greater than zero (non-negativity condition).

$$\begin{aligned}
\text{Objective function } (\min \textit{ or } \max) \quad & Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n && (2.1)\\
\text{subject to} \quad & && (2.2)\\
& a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \ (\leq, =, \geq)\ b_1 && (2.3)\\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \ (\leq, =, \geq)\ b_2 && (2.4)\\
& \quad \vdots \qquad\quad \vdots \qquad\quad \vdots && \\
& a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \ (\leq, =, \geq)\ b_m && (2.5)\\
\text{with} \quad & x_1,\ x_2, \ldots,\ x_n \geq 0 && (2.6)
\end{aligned}$$

Dantzig (1990) proposed the simplex method, used to solve general LP problems. Thus, when we consider n decision variables, the simplex method has exponential performance and visits $O(2^n)$ vertices in the worst-case scenario (Klee and Minty, 1972), but Spielman and Teng (2004) showed that the algorithm had expected polynomial behavior for most of the real-world linear problems.

In mathematics, a simplex is an object connecting $n+1$ points in an n-dimensional space, if we consider the dimension $D = 1$, the simplex is a line segment connecting 2 points (Sharma, 2017, ch 4, p. 101). Considering the LP defined in (2.1)—(2.6), and its linear problem standard form definition (2.7)—(2.9), the simplex method can find global optima by iterating between a finite set of vertices, provided by the intersection of the problem's constrains (2.8).

$$\begin{aligned}
\max \quad & Z = c^T x && (2.7)\\
\text{subject to} \quad & Ax = b && (2.8)\\
& x \geq 0 && (2.9)
\end{aligned}$$

The gray area in Figure 2.5 represents the problem's feasible region, and the intersection red dots are the primary feasible solution vertices. In each iteration step, the algorithm evaluates the objective function (2.7) according to the current vertex and then visits the next one until it finds the optimal basic feasible solution.

Through Integer Linear Programming (ILP), we can model and solve integer linear problems, which belong to the $\mathcal{NP}$-hard problem class. A typical integer linear problem has the standard form of Equation (2.10)–(2.13), and as established by (2.13), each of its variables have integer

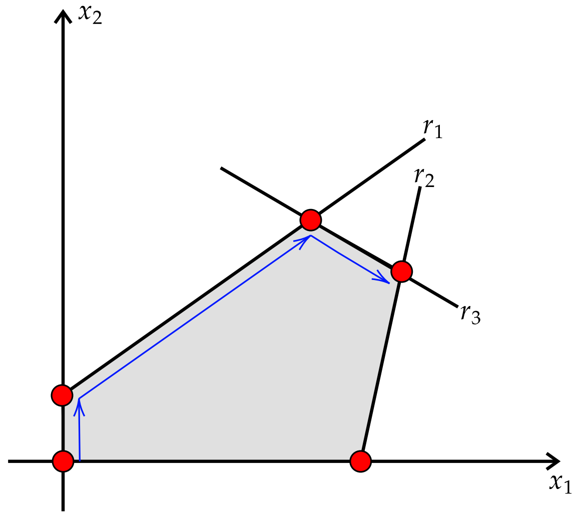


Figure 2.5: The simplex execution process.

values. There are classic and well-established algorithms to solve integer linear problems, such as the branch and bound (Lawler and Wood, 1966), branch and cut (Padberg and Rinaldi, 1987), and the cutting-plane method (Kelley, 1960).

$$\max \quad Z = c^T x \tag{2.10}$$

$$\text{subject to} \quad Ax = b \tag{2.11}$$

$$x \geq 0 \tag{2.12}$$

$$x \in \mathbb{Z} \tag{2.13}$$

2.3.2 Branch and bound

The Branch and Bound algorithm is an exact method for solving integer problems, proposed by Land and Doig (2010). It has application in many classic optimization problems, such as the traveling salesman problem (Applegate et al., 2006), the generalized assignment problem (Ross and Soland, 1975), and the knapsack problem (Kolesar, 1967). The branch and bound algorithm organizes the search method as a tree. Thus, it creates a source node split into subproblems connected to the source. As an example, we can consider the general linear problem (2.14) (a simplified form of Equation (2.1)), in which x is a decision variable vector and the set of constraints presented in Expression (2.15).

$$\max \quad c^T x \tag{2.14}$$
$$\text{subject to} \quad Ax \leq b \tag{2.15}$$
$$x \in \mathbb{Z} \tag{2.16}$$

The method starts by solving the linear programming relaxation of an instance of the problem defined in Equation (2.14)–(2.15), i.e., we remove the integrality constraint of each variable. We test if the solution x^* is composed of integer values. If x^* contains any fractional value, it is not a valid solution, and the algorithm selects a non-integer variable $x_j^* \in x^*$ to generate two new variables $\beta^{'}$ (floor, greatest integer less than x_j^*) and $\beta^{''}$ (ceiling, the smallest integer greater than x_j^*), as established by Equation (2.17). We use these values to create the new subproblem's branches, as defined by Equations (2.18)—(2.19).

$$(\beta^{'}, \beta^{''}) = (\lfloor x_j^* \rfloor, \lceil x_j^* \rceil) \tag{2.17}$$
$$\max \quad c^T x \quad \text{subject to} \quad Ax \leq \beta^{'},\ x\ in\ S \tag{2.18}$$
$$\max \quad c^T x \quad \text{subject to} \quad Ax \geq \beta^{''},\ x\ in\ S \tag{2.19}$$

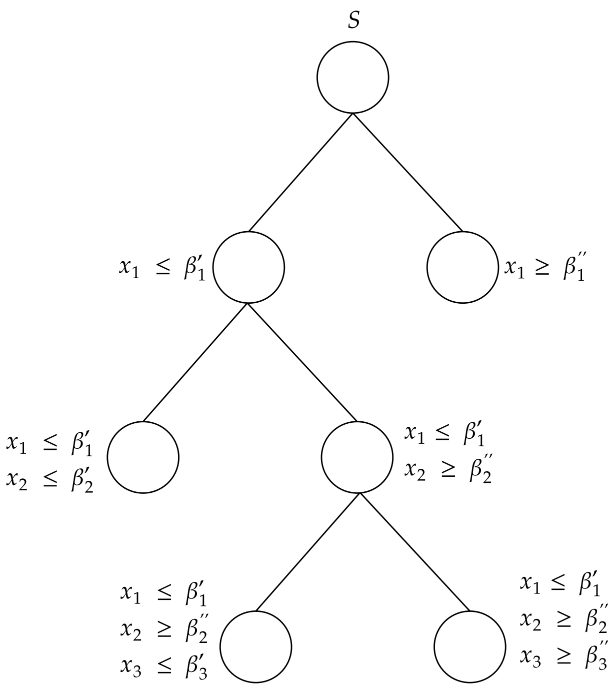


Figure 2.6: The branch and bound tree (Applegate et al., 2006).

In Figure 2.6, problem (2.14) represents the root node as initial state of the algorithm, followed by the branching subproblems (2.18) and (2.19). Each branch and bound tree node has a value for the objective function, according to its x variables. The algorithm decides if it is worth expanding new nodes by evaluating the objective function value and checking if x^* contains any fractional value. We achieve a final solution if there is no feasible solution or if x^* for a given tree node is integer only and contains the best value for the objective function. We use the branch and bound algorithm to find optimal solutions for the simulated mobility scenarios.

2.3.3 Iterated local search

The Iterated Local Search (ILS) is a sequence of solutions built iteratively by an embedded heuristic. (Lourenço et al., 2019). In its history, it received many names and propositions (Baxter, 1981; Martin et al., 1991; Martin and Otto, 1996), and the demand for new approaches for different surging problems continues to push the efficiency boundaries through evolutionary optimization (Cao et al., 2018).

Algorithm 2.1 describes the general framework of the iterated local search. The procedure is simple and starts by generating the initial solution S_0 from the starting instance I. In the following step, the algorithm performs a local search to obtain a new solution S, stores it in S^*, which is an improvement S_0. Thus, the local search method aims to find better solutions by making calculated adjustments in the current input instance.

Algorithm 2.1 General iterated local search

procedure ILS($I, maxIter$)
    $S_0 \leftarrow$`Initialize`(I)
    $S \leftarrow$`Local_Search`(S_0)
    $S^* \leftarrow S$
    $iter \leftarrow 0$
    while $iter < maxIter$ **do**
        $S' \leftarrow$`Perturbate`(S)
        $S' \leftarrow$`Local_Search`(S')
        `Accept`(S, S')
        `Update_Best_Solution`(S, S^*)
        $iter \leftarrow iter + 1$
    end while
    return S^*
end procedure

The loop structure updates the number of performed iterations $iter$ to avoid exceeding the maximum limit $maxIter$. The iterations repeat a series of perturbations, local searches, and cost checks to provide the best result as the final output. Since the iterative process is prone to get trapped in local maximum or minimum values, the perturbation function `Perturbate` can randomly change the current solution variables. As the expected result 2.7, the algorithm steers the solution away from local optima. The acceptance function `Accept`$(S, S^{'})$ compares S and $S^{'}$, allowing that the function `Update_Best_Solution`(S, S^*) update their current values to the new best if $cost(S^{'}) < cost(S)$.

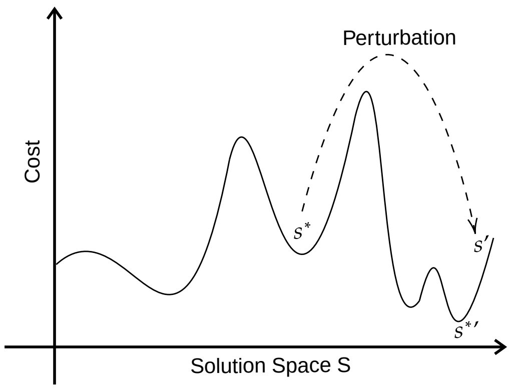


Figure 2.7: Expected ILS perturbation result.

2.3.4 Variable neighborhood descent

The Variable Neighborhood Descent (VND) (Mladenović and Hansen, 1997) is one of the most used and efficient local search algorithms. In the context of our work, we implement the `Local_Search`(S_0) with the VND Algorithm 2.2, which seeks the best solution through a search of sorted neighborhood sets. Considering $V = \{V^1, \ldots, V^r\}$ a set of neighborhoods of the graph G, the VND starts V. Starting from $k = 1$ until a max of k_{max} iterations, the algorithm looks for neighborhood sets that reduce a defined cost g^* of the graph G. As long as it is possible to find a neighborhood G' with a lower cost, which we will consider as the solution S, the algorithm will try to find better neighborhoods, and if not possible, we move on to the next neighborhood $k+1$. We will use the VND as the local search of the ILS Algorithm 3.1, this combination presents successful cases in the literature (Penna et al., 2011).

Algorithm 2.2 General variable neighborhood descent

procedure VND(G)
    start V
    $k \leftarrow 1$
    $g^* \leftarrow Cost(G)$
    repeat
        $G' \leftarrow V^k(G)$
        if $Cost(G') < g^*$ **then**
            $S \leftarrow G'$
            $k \leftarrow 1$
        else
            $k \leftarrow k+1$
        end if
    until $k = k_{max}$
    return S
end procedure

2.3.5 Generalized assignment problem

We will reserve some time to discuss the GAP (Generalized Assignment Problem) (Cattrysse and Van Wassenhove, 1992), as shown in Figure 2.8, as our proposition can be defined as one of its specific case. This problem is part of the $\mathcal{NP}$-hard complexity class (Sahni and Gonzalez, 1976), and is generally defined according to Equation (2.20). The value of c_{ij} represents the cost of assigning a job $j \in J$ to the agent $i \in I$. The variable x_{ij} indicates if i receives the job j, and in this case $x_{ij} = 1$. If the job j is not assigned to the agent, then $x_{ij} = 0$. In Figure 2.8, which is a complete bipartite graph, the jobs $J = \{1,2,...,m\}$ are assigned to the agents $I = \{1,2,...,n\}$ according to the blue colored edges. Each job is assigned to at least one agent, and there is a occurrence of one agent ($i = 1$) with more than one job.

$$(GAP)\ \min \quad \sum_i \sum_j c_{ij} x_{ij} \tag{2.20}$$

$$\text{subject to} \quad \sum_j a_{ij} x_{ij} \leq b_i \qquad \forall\, i \in I \tag{2.21}$$

$$\sum_i x_{ij} = 1 \qquad \forall\, j \in J \tag{2.22}$$

$$x_{ij} \in \mathbb{B} \qquad \forall\, i \in I,\ \forall\, j \in J \tag{2.23}$$

The assignments aim to minimize the overall costs of an objective function by performing multiple combinations. The solution must satisfy conditions such as the Knapsack (Salkin and De Kluyver, 1975) set of constraints (2.21), and the restriction that each job must be assigned for at least one agent in this particular case (2.22). Each agent has an individual level of effort or capacity a_{ij} to offer, and if a job exceeds this effort level, we must assign it to another agent. Considering the set of agents I, the sum of their capacity must not exceed the established max capacity b_i.

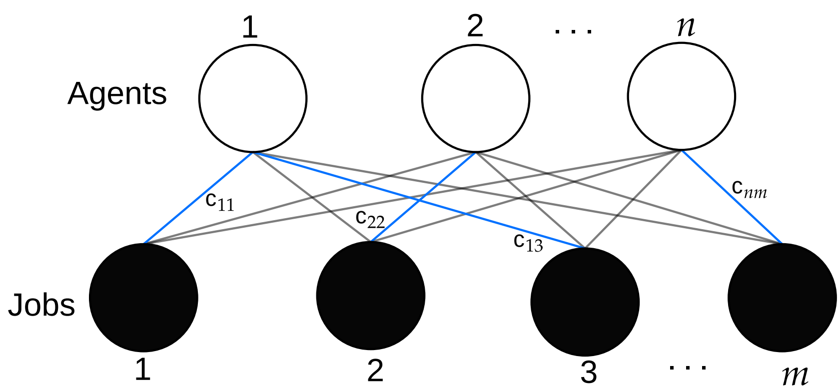


Figure 2.8: Attribution of jobs to agents according to costs, a general assignment problem.

In real-life scenarios, the GAP appears in different ways, such as creating efficient schedules for workers or building timetables for teachers and classes. It also has variations such as the multilevel, the nonlinear capacity constrained, and the bottleneck GAP (Öncan, 2007). Literature has extensive documentation and solution, and the most classical approaches propose approximation methods (Shmoys and Tardos, 1993), and lower/upper bound manipulation through the branch and bound algorithm (Ross and Soland, 1975). Modern approaches usually propose minor changes to the classic algorithms, hybrid solutions, or genetic and evolutionary based-solutions (Srivarapongse and Pijitbanjong, 2019; Jia et al., 2018).

We will see in Chapter 3 that our allocation models, defined in Equations (3.8) and (3.19), have the same standard form of the GAP (2.20). Similarly to assigning jobs to agents, our model allocates (assigns) users to base stations according to a set of constraints. Each user must connect to at least one base station, and each base station can receive as many users as its maximum bandwidth capacity allows.

This chapter introduced key concepts related to this work. In Section 2.1, we discussed the advances in mobile networks technologies, which are mainly transitioning to the fifth-generation. By allocating users in base stations, our contribution impacts mobile networks and their handover processes mentioned in Section 2.2. In regards to optimization, our solution can be deployed by using some of the exact or heuristic methods discussed in Section 2.3.

3

METHODOLOGY

In this chapter, we introduce Taleb et al. (2015) optimization model in Section 3.1. The authors propose a multi-objective function to minimize handovers between eNBs, and the communication cost between eNBs and data centers. Their work is a starting point for our proposal, which extends their model. In Section 3.2, we present our contribution by first showing a distance-based allocation model, followed by the heuristic solution 3.2. We change to a heuristic approach as an alternative to the exact method and consider the RSRQ (Reference Signal Received Quality) indicator instead of distance. The RSRQ informs the quality of the wireless communication between UEs and eNBs.

3.1 Handover and communication cost optimization

Based on the Multi-Objective model of Taleb et al. (2015), our proposal aims to improve the mobility and connection in wireless networks. The authors consider the EPS (Evolved Packet System) architecture (Figure 3.1). The S-GW (Serving Gateway) is responsible for routing packages, mobility management, and handover processes. The HSS (Home Subscriber Server) saves identification and location of UEs, and it also has authentication and authorization functionalities. The PDN-GW (Packet Data Network Gateway) bridges networks (the Internet) containing several data centers and entities such as the eNBs and UEs. The MME (Mobile Management Entity) contains UE's mobility management functionalities and location information, interacting with the HSS entity to execute authentication. The eNBs connect UEs with the rest of the network.

Taleb et al. (2015) work establishes a model to minimize the communication cost between eNBs and data centers and a model that minimizes UE's handover between coverage areas. These two models compose a multi-objective optimization problem. The authors consider an SDN context with virtual network infrastructure, in which the S-GW and the PDN-GW have virtual

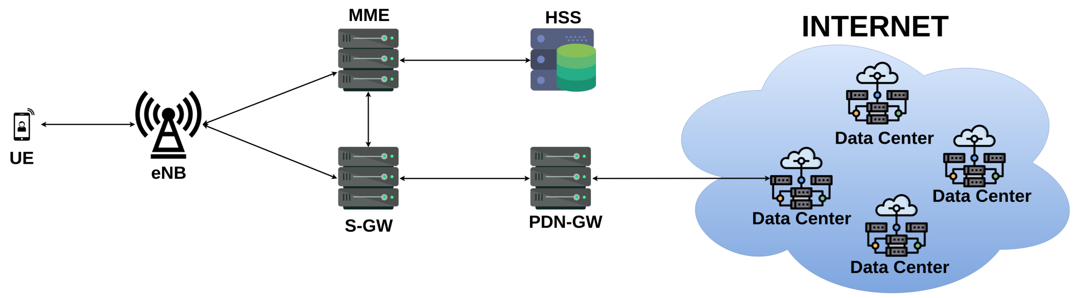


Figure 3.1: EPS architecture for LTE networks.

function sets allocated for distinct entities depending on the user's demand and behavior. In the communication cost minimization problem, a solution is to allocate the virtual functionalities of the PDN in the data centers of the nearby eNBs. To minimize handovers, the authors place virtual functions of S-GW in more distant regions, creating more significant coverage areas.

Considering a set of eNBs $= \{i_1, i_2, ..., i_{N-1}, i_N\}$, where N is the total base stations, and a set of data centers $= \{s_1, s_2, ..., s_{DC-1}, s_{DC}\}$ containing a total of DC data centers, Taleb et al. (2015) complete model is defined as follows:

$$
\begin{aligned}
\min \quad & (f(h_{ij}, x_{ij}), g(c_{is}, y_{is})) && && (3.1)\\
\text{subject to} \quad & y_{is} + y_{js} \leq 1 + x_{ij} && \forall\, i, j \in N,\ \forall\, s \in DC && (3.2)\\
& y_{is} - y_{js} \leq 1 - x_{ij} && \forall\, i, j \in N,\ \forall\, s \in DC && (3.3)\\
& \sum_{s \in DC} y_{is} = 1 && \forall\, i \in N && (3.4)\\
& y_{is}, x_{ij} \in \mathbb{B} && \forall\, i, j \in N,\ \forall\, s \in DC && (3.5)
\end{aligned}
$$

The y_{js} variable is equal to 1 if eNB$_j$ is connected to the data center s, and 0 otherwise. The multi-objective function (3.1) uses two criteria f and g. The constraints (3.2) states that two eNBs must not connect to the same data center, which means $(x_{ij} = 0) \implies (y_{is} = 0) \vee (y_{js} = 0)$. Constraints (3.3) states that two connected eNBs must not be connected to different data centers, which implies $(x_{ij} = 1) \implies (y_{is} = y_{js})$. According to (3.4), every eNB must be connected to only one data center, and the variables y_{is} and x_{ij} belong to the binary domain, as we can see in (3.5).

The objective functions f e g are respectively defined by (3.6) for handover and (3.7) for cost. The values of h$_{ij}$ represent the average handover frequency between eNB$_i$ and eNB$_j$. Base stations connected to the same data center belong to the same coverage area, and in such cases, there is no handover between the eNBs. The values of c_{is} represent the communication cost between the eNB$_i$ and the data center s. The purpose of the objective function (3.7) is to select an eNBs$_i$ near a data centers s of the set DC, assuring the smaller communication cost.

$$f(h_{ij}, x_{ij}) = \min \sum_{i \in N} \sum_{j \in N} h_{ij}(1 - x_{ij}) \tag{3.6}$$

$$g(c_{is}, y_{is}) = \min \sum_{i \in N} \sum_{s \in DC} c_{is} y_{is} \tag{3.7}$$

Since Taleb et al. (2015) model (3.1) did not consider the allocation of UEs in eNBs, our proposal extends the current model by implementing it. The proposed model allocates users in base stations by using distance, average handover frequency for each eNB, and the bandwidth requirements of UEs and eNBs. It is worth mentioning that the handover and cost criteria f and g have opposite natures since to decrease handover between coverage areas, we need to increase the distance between eNBs and data centers to create more significant coverage areas. On the other hand, we need to reduce this distance to reduce the communication cost between eNBs and data centers. As a consequence, we need to find the ideal exchange between these two objectives. We can use multi-objective optimization approaches (Marler and Arora, 2004).

3.2 Proposed model

We introduce the allocation of UEs in eNBs, which can occur according to available bandwidth, distance, and each eNBs' handover average. Figure 3.2 illustrates an example that is not covered in Taleb et al. (2015). In this example scenario, composed of 5 eNBs, horizontally spaced by 50 units of distance, consider the max bandwidth capacity L_i = 20 Mbps for each eNB$_i$. Each user UE$_k$ has a minimum bandwidth requirement l_k = 3 Mbps. Therefore, each eNBs can not support service to more than six users simultaneously. We recalculate the solution dynamically when the UE moves between eNBs.

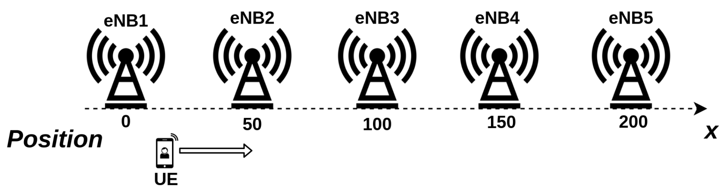


Figure 3.2: An example of an UE moving along a set of eNBs.

For such a scenario, we can apply the new objective function defined by (3.8). Notice that our allocation model has the GAP (2.20) standard form. We will use this property to evaluate the results of the heuristic method presented in Subsection 3.2. The variable b_{ki}, if its value is equal to 1, maps a UE$_k$ that is connected to an eNB$_i$, and if they are not connected, b_{ki} = 0. Equation (3.8) establishes that the users connect to nearby base stations with the smallest handover average h_i.

$$\min \quad z(b_{ki}, d_{ki}, \overline{h_i}) = \sum_{k \in U} \sum_{i \in N} b_{ki}(d_{ki} + \overline{h_i}) \tag{3.8}$$

Equation (3.9) gives the formula for calculating h_i, which can be found with the sum of the average handover frequency h_{ij} for each eNBs$_j \in N$, connected to eNB$_i$, divided by $N-1$ (we remove eNB$_i$ from N). The value of d_{ki} represents the distance between UE$_k$ and eNB$_i$.

$$\overline{h_i} = \sum_{j \in N \setminus \{i\}} \frac{h_{ij}}{(N-1)} \qquad \forall\, i \in N \tag{3.9}$$

The model constraints are given by (3.10) — (3.12). We establish in (3.10) that every user must be connected to only one base station. By constraints (3.11), the max bandwidth capacity L_i of a base station i must not exceed the sum of the bandwidth minimum requirements l_k for each user $k \in U$. The binary domain of b_{ki} if defined in (3.12).

$$\sum_{i \in N} b_{ki} = 1 \qquad \forall\, k \in U \tag{3.10}$$

$$\sum_{k \in U} l_k b_{ki} \leq L_i \qquad \forall\, i \in N \tag{3.11}$$

$$b_{ki} \in \mathbb{B} \qquad \forall\, i \in N \tag{3.12}$$

To find the value of d_{ki} in a Cartesian coordinate systems, we use Equation (3.13). Therefore, $d_{ki} = d_{cart}$ is the Euclidean distance between UE$_k$ at position $p_k = (x_k, y_k, z_k)$, and eNB$_i$ at position $p_i = (x_i, y_i, z_i)$. To geographic coordinate systems, $d_{ki} = d_{harv}$, according to the Haversine formula (3.14) to spherical surfaces (Sinnott, 1984). In this case, (ψ_i, β_i) represents respectively eNB$_i$'s latitude and longitude. In the same way, (ψ_k, β_k) is UE$_k$'s latitude and longitude. The Earth is our sphere, and R is its radius, which is approximately 6317 Km.

$$d_{cart} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2} \tag{3.13}$$

$$d_{harv} = (2R) \arcsin\left(\sqrt{\sin^2\left(\frac{\psi_i - \psi_k}{2}\right) + \cos(\psi_k)\cos(\psi_i)\sin^2\left(\frac{\beta_i - \beta_k}{2}\right)}\right) \tag{3.14}$$

Equations of handover (3.6), cost (3.7), and user allocation (3.8) are used as criteria in a Multi-Objective Function (COMP). We will use the weighted sum method described by Equation (3.16) and (3.17). This approach provides a multi to mono-objective conversion for a set with m

optimization criteria, in which each F_i criterion will be weighted by a value $w_i \in [0,1]$. The sum of all the criterion weights must be equal to 1. Varying the value of w_i means assigning different importance to a specific objective function. If w_2 = 0.6, then $g(c_{is}, y_{is})$ has a greater contribution for worsening the optimal solution of (COMP). As a consequence, smaller values of c_{is} will be required, which means that we are prioritizing communication cost minimization. The complete proposed formulation is given by:

$$\begin{aligned} \text{(COMP)}: \min \quad & obj = w_1 \times f(h_{ij}, x_{ij}) + w_2 \times g(c_{is}, y_{is}) + w_3 \times z(b_{ki}, d_{ki}, \overline{h_i}) \\ \text{subject to} \quad & (3.2-3.5) \text{ and } (3.10-3.12) \end{aligned} \tag{3.15}$$

$$\sum_{i=1}^{m} w_i F_i(x) \tag{3.16}$$

$$\sum_{i=1}^{m} w_i = 1 \tag{3.17}$$

We could execute the complete formulation (COMP) in network entities such as the mobile manage entity presented in Figure 3.1. If we consider an SDN context, a controller executes the model and distributes virtual network functions of PDN-GW and S-GW according to results. This strategy should create coverage areas with reduced handovers, minimum communication cost between eNBs and data centers, and the allocation of UEs in the best candidate eNB. We must connect every user to one base station, keeping network bandwidth limits. When executed, the model must run periodically, collecting and updating information about eNBs and UEs.

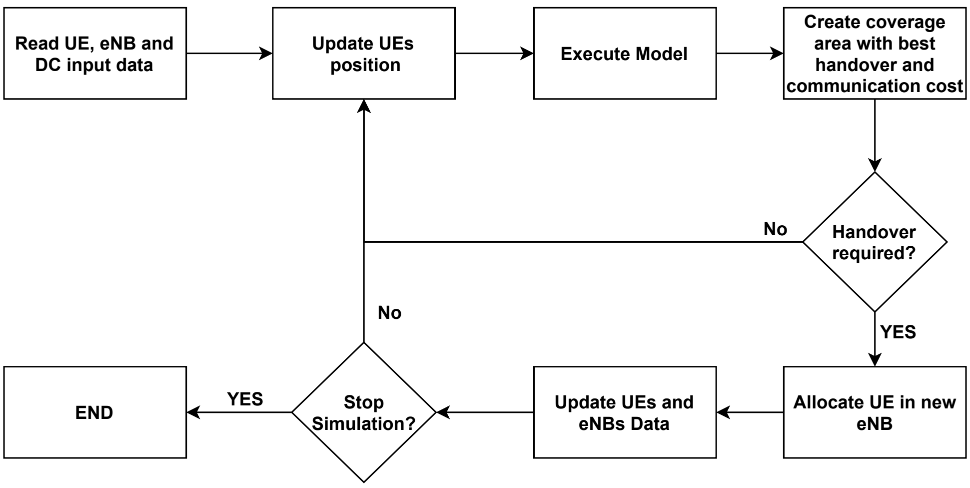


Figure 3.3: Simulation flowchart.

We can implement this model in the linear programming solver CPLEX (IBM, 2017), which uses the branch and bound exact method. The simulation process follows Figure 3.3, starting by reading the eNBs' handover average, position, and bandwidth requirements for UEs and eNBs. It proceeds with reading updates of UEs' position and execution of (COMP).

As a result, (COMP) provides a set of coverage areas connecting the eNBs with minimum handover frequency. These eNBs connect to the data centers with lower communication costs. If handover is required, the model allocates users to a new eNB and proceeds to update users' and eNBs data. If the simulation does not end, we read UE's position data to restart the process.

Heuristic approach

Our previous user allocation model defined in Equation (3.8) allocates UEs based on the distance and the eNB individual handover frequency average. However, due to noise and physical environment interference, the distance may not be the best indicator for connection quality. We can improve this model by considering indicators such as the Received Signal Strength Indication (RSSI), the Reference Signal Received Quality (RSRQ), or the Reference Signal Received Power (RSRP).

The RSSI is an indicator of wireless signal strength. The RSRP presents the base station average signal power perceived by the users. Finally, the RSRQ informs the received signals' quality. It is a commonly used parameter in handover solutions since it contains information on noise and interference levels. We will simulate the RSRQ value according to Equation (3.18), in which the signal degrades with distance. For example, if we consider the RSRQ values in Table 3.1, the maximum transmission radius R_i of a given base station i, and the distance $d_{ki} = R_i$ for the user k, then $\Theta_{ki} = -12$ dB, which is a poor RSRQ value ($d_{ki} = 0 \implies \Theta_{ki} = -5$ dB).

$$\Theta_{ki} = -[\frac{d_{ki}}{R_i}(|\Theta_{min}| - |\Theta_{max}|) + |\Theta_{max}|] \tag{3.18}$$

Table 3.1 presents the signals' quality classification according to their respective values. In ideal conditions, users with poor signal reception are further away from the eNB or located at cell edges. Excellent or good signal reception means that the user is closer to the base station.

Table 3.1: LTE signal quality indicators.

Signal Quality	RSSI (dBm)	RSRQ (dB)	RSRP (dB)
Excellent	-65	-5	-84
Good	-65 to -75	-9 to -5	-85 to -102
Fair	-75 to -85	-12 to -9	-103 to -111
Poor	-85	-12	-111

We will reformulate the user allocation model (3.8) to consider the average handover frequency h_i of each eNB, and the RSRQ value Θ_{ki}, which represents the communication quality between the user k and eNB$_i$. The model p in Equation (3.19) allocates users in the eNBs with the minor handover frequency average h_i and with the best signal quality Θ_{ki}.

$$\min \quad p(b_{ki}, \overline{h_i}, \Theta_{ki}) = \sum_{k \in U} \sum_{i \in N} b_{ki}(\overline{h_i} + \Theta_{ki}) \tag{3.19}$$
$$\text{subject to} \quad (3.10 - 3.12)$$

We will not consider Taleb et al. (2015) models in the next evaluations, but similarly to (COMP), we can use the solving constraints and weighted sum method to define (COMP2).

$$\text{(COMP2)}: \min \quad obj = w_1 \times f(h_{ij}, x_{ij}) + w_2 \times g(c_{is}, y_{is}) + w_3 \times p(b_{ki}, \overline{h_i}, \Theta_{ki}) \tag{3.20}$$
$$\text{subject to} \quad (3.2 - 3.5) \text{ and } (3.10 - 3.12)$$

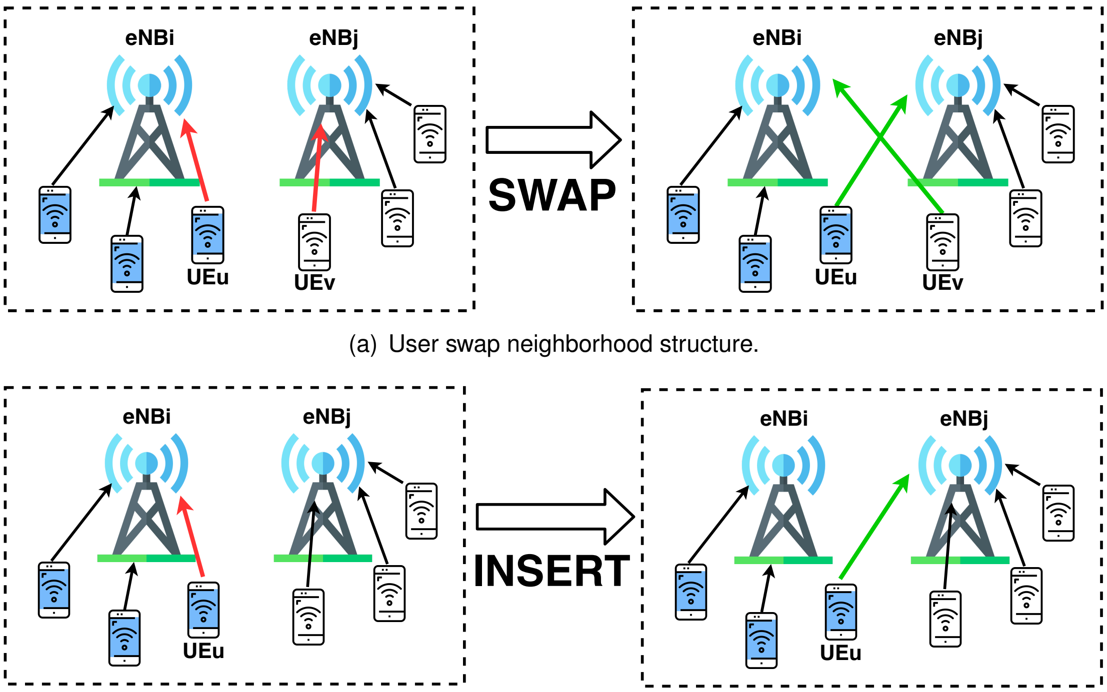
We will evaluate heuristic methods to solve our allocation model as an alternative to the branch and bound exact algorithm. Heuristic approaches do not guarantee optimal results. However, there are scenarios in which the optimal solution exceeds the time limit or computational resources are limited, so we have to use a solution sufficiently good that does not exceed hardware, software, or time limitations.

Our heuristic method uses the iterated local search (ILS) metaheuristic (Lourenço et al., 2019) coupled with the variable neighborhood descent (VND) (Hansen et al., 2018) in the local search phase. Also, to complement the search algorithms, two neighborhood structures are defined to execute swap and insertion operations: *swap_ue(UE$_{ui}$, UE$_{vj}$)*, and *insert_ue(UE$_{ui}$, eNB$_j$)*. These structures can only be executed if the model constraints are obeyed.

- *swap_ue(UE$_{ui}$, UE$_{vj}$)*: This neighborhood structure swaps a user UE$_u$, allocated in eNB$_i$, with a user UE$_v$, allocated in eNB$_j$, as shown in Figure 3.4(a). This procedure stops the local search at the first swap that improves the current solution. It has a worst case complexity $O(n^2)$ for a instance with n users.

- *insert_ue(UE$_{ui}$, eNB$_j$)*: This neighborhood structure removes a user UE$_u$, allocated in eNB$_i$, and inserts (or allocates) it in eNB$_j$. Figure 3.4(b) illustrates an example of a insert operation between two eNBs. This procedure stops the local search at the first insertion that improves the current solution. It has a worst case complexity $O(nm)$ for a instance with n UEs and m eNBs .

The cost improvement evaluations of the insertion and swap neighborhood structures take $O(1)$ complexity. Algorithm 3.1 presents the iterated local search. Starting from a graph G that will represent our network, an initial solution S_0 is generated with a greedy algorithm (Cormen



(a) User swap neighborhood structure.

(b) User insert neighborhood structure.

Figure 3.4: The two neighborhood structures for swap and insertion operations.

et al., 2009, chapter 16), implemented in the function $GenerateInitialSolution(G)$. This function randomly selects UEs and allocates them in the eNB with minimum cost. The initial solution S_0 represents users connected to eNBs. Upon S_0, we use the VND based local search Algorithm 3.2 to find a new solution, denoted by S.

Algorithm 3.1 Iterated local search

procedure ILS(G)
    $S_0 \leftarrow$ GenerateInitialSolution(G)
    $S \leftarrow$ LocalSearch(S_0)
    $counter \leftarrow 0$
    while !(stop condition) **do**
        $S' \leftarrow$ `Perturbation`(S)
        $S' \leftarrow$ `LocalSearch`(S')
        $counter \leftarrow counter + 1$
        if $Cost(S') \leq Cost(S)$ **then**
            $S \leftarrow S'$
        else if $counter \geq n$ **then**
            $counter \leftarrow 0$
            $S' \leftarrow$ GenerateInitialSolution(G)
        end if
    end while
    return S
end procedure

The VND starts by setting the improvement index k and its maximum value k_{max}. There is an $improvement$ variable to indicate if any of the neighborhood structures found improvement (returning true) at the current iteration and a flag that resets the k value in this situation. To prevent the algorithm from running indefinitely, we increment k each time a neighborhood structure improvement fails (returning false). If the two neighborhood improvements fail, then k = 3, and the VND algorithm ends.

Algorithm 3.2 Variable neighborhood descent

procedure VND
  $k \leftarrow 1$
  $k_{max} \leftarrow 3$
  $improvement \leftarrow false$
  $flag \leftarrow false$
  repeat /*eNBs and UEs randomly chosen*/
    if $k = 1$ **then**
      $flag = insert\_ue(UE_{ui}, UE_{vj})$
    end if
    if $k = 2$ **then**
      $flag = swap\_ue(UE_{ui}, eNB_j)$
    end if
    if (flag) **then** /*Reset k if any neighborhood structure returns true (improvement found)*/
      $k \leftarrow 1$
    else
      $k \leftarrow k + 1$
    end if
  until $k = k_{max}$
  return $improvement$
end procedure

The local search generates different solutions by performing the insert and swap operations with the mentioned neighborhood structures. Subsequently, we enter the loop of the iterated local search, which contains a perturbation function to avoid local minimums. Finally, from the disturbed solution S', we execute a local search again. Considering that the global costs of the networks S and S', if the condition $Cost(S') \leq Cost(S)$ is satisfied, then S' is assigned to S as a new solution. The final result is a minimized cost in the relationships between users and base stations.

In terms of allocation and handover, we can compare our proposed model of Equation (3.19) with the solution proposed by Ahmadi et al. (2020). To allocate users, their model builds a base station score rank, in which each eNB receives a score according to Equation (3.21). From (3.22), if a user is approaching a given eNB and simultaneously the RSRQ value is improving, then $factor_{direction} = 1$, and 0 otherwise. The $factor_{distratio}$ depends on the user's current position and all possible future routes. The eNBs which cover most future routes receive the highest scores. As for the RSRQ parameter, the better the signal, the better the score. All values are normalized (0 to 1).

$$score = RSRQ + Factor_{prediction} \tag{3.21}$$

$$Factor_{prediction} = factor_{direction} + factor_{distratio} \tag{3.22}$$

This chapter mainly discussed this work's mathematical models. Taleb et al. (2015) propose a handover and communication cost minimization model, which we discussed in Section 3.1. As a contribution, we proposed an allocation model to address the relation between users and base stations in Section 3.2. We refined our model by considering parameters such as the RSRQ signal indicator and providing a heuristic approach.

4

RESULTS AND DISCUSSION

This chapter presents the computational experiments conducted to evaluate the effectiveness of our proposed approaches. The experiments are divided into three sections. First, in Section 4.1, we analyze the performance of the new ILP formulation applied in the simulation process described in Figure 3.3. Second, in Section 4.2 we evaluate the quality of the solutions found by the ILS metaheuristic. Finally, in Section 4.3, we conduct an extensive simulation, and compare our proposal with Ahmadi et al. (2020) approach, considering allocation (RSRQ impact) and handover.

4.1 Performance of ILP formulation

This experiment evaluates the proposed distance-based model (COMP), implemented with the IBM ILOG CPLEX Optimization Studio v12.5.0 solver and executed on Intel Core i5-6200U (2.30GHz and 4 cores) with 4GB of RAM, Ubuntu 16.04 LTS operating system.

Execution time

Table 4.1 shows the average execution time for different combinations of the weights w_1, w_2 and w_3. For this test case, we considered 10 data centers and a sample of 60 base stations in the city of Maceió, Alagoas, available in Telebrasil (2021) database, which has latitude and longitude data of several base stations of different mobile carriers in Brazil. Ten users are randomly placed and maintained in positions near eNBs so that the model can calculate distances. Since handover data and communication costs between data centers and eNBs are not available, we considered a random uniform distribution to generated 30 instances of these values for each eNB. We compute the execution time for every instance, and in the end, we calculate the average execution time for the different weight combinations.

Table 4.1: Average execution time for different weights combinations. Total number of instances = 30, eNBs = 60, data centers = 10, UEs = 10.

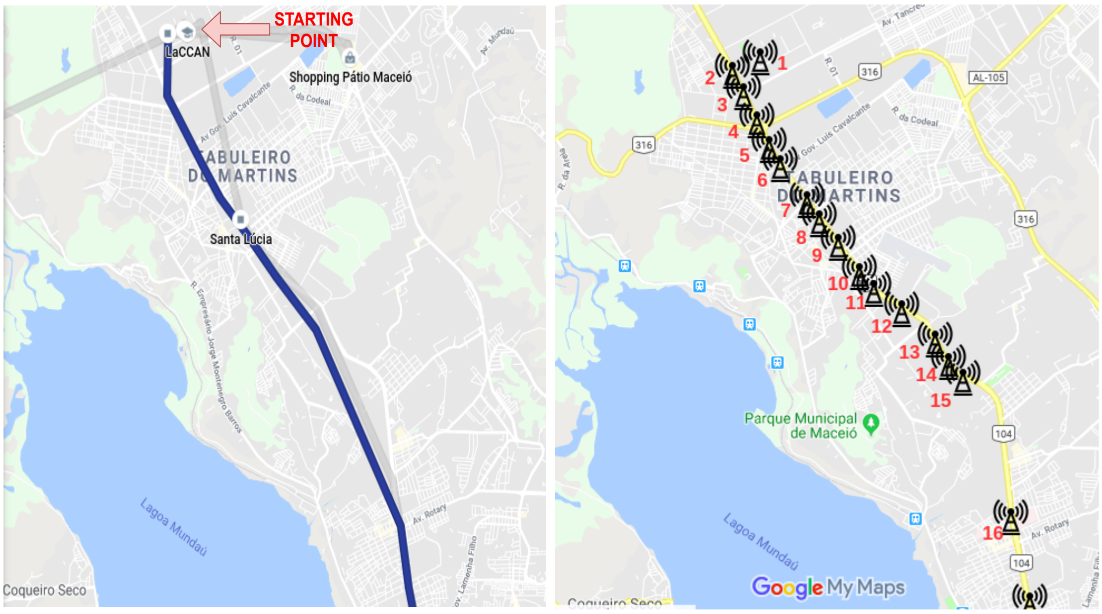
w_1	w_2	w_3	Average execution time (seconds)
1	0	0	0.75
0	1	0	1.09
0	0	1	1.13
0.8	0.1	0.1	2.04
0.1	0.8	0.1	4.44
0.1	0.1	0.8	2.42
0.4	0.3	0.3	2.69
0.3	0.4	0.3	2.70
0.3	0.3	0.4	2.42
0.9	0.1	0	1.69
0.1	0.9	0	4.17

In the the formulation (COMP), the weight configurations $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ verify the execution time in an isolated manner for $f(x_{ij},h_{ij})$, $g(c_{is},y_{is})$ and $z(b_{ki},d_{ki},\overline{h_i})$, since only one of the three criteria is considered. In a case of assigning higher priority to a specific criterion, it is observable that the execution time increases considerably for the weights combination $(0.1,0.8,0.1)$, which is when the costs of communication between eNBs and data centers need to be strictly optimized. If an approximately uniform combination such as $(0.4,0.3,0.3)$, $(0.3,0.4,0.3)$ or $(0.3,0.3,0.4)$ is to be taken, the average execution times are similar. We can verify Taleb et al. (2015) base model average execution time by making $z(b_{ki},d_{ki},\overline{h_i}) = 0$, which results in Equation (3.1). Configuration $(0.1,0.9,0)$ shows that it takes longer to find an optimized solution to minimized communication costs with high priority, and low priority for reduced number of handover between coverage areas. In the opposite case, $(0.9,0.1,0)$, a considerable smaller average execution time is required.

User allocation

We collected latitude and longitude data from the Google Maps location history of a moving UE. Figure 4.1(a) shows the starting point of a 13.7 km displacement, traversed in a time interval of 42 minutes. The user followed the path from the starting point to the highest-numbered base station (21 in total), some of them illustrated by Figure 4.1(b). In addition, we found several base station locations in the database provided by Telebrasil (2021). In this study case, for every new UE's location, the model must evaluate if it is worth keeping it allocated in the same eNB or if it is better to transfer it to another one. Haversine Equation (3.14) gives the distance between user and base station in this scenario.

Figure 4.2 shows the eNBs where the system allocates the UE over the 13.7 km path. According to its GPS information, the initial UE position was nearby eNB$_1$ and eNB$_2$, with respective distances of 0.08 km and 0.45 km. The average handover of these base stations is 5.3 for the



(a) eNBs placement along the path traveled by UE. (b) eNBs in which the UE was allocated over the path.

Figure 4.1: Path traveled by UE over time, moving forward from the starting point to the highest-numbered eNB.

first and 3.75 for the second. The formulation (COMP) allocates UEs in eNBs with the smallest sum of distance and handover average. Thus, it is observable in Figure 4.2 the correct allocation in eNB$_2$. During the last 10 minutes of the traveled path, eNB$_{19}$, eNB$_{20}$ and eNB$_{21}$ are closer, at a respective distance of 0.78, 0.91, and 0.53 kilometers. Their handover averages are 3.8, 6.25, 5.5, which causes the permanence of UE in eNB$_{19}$.

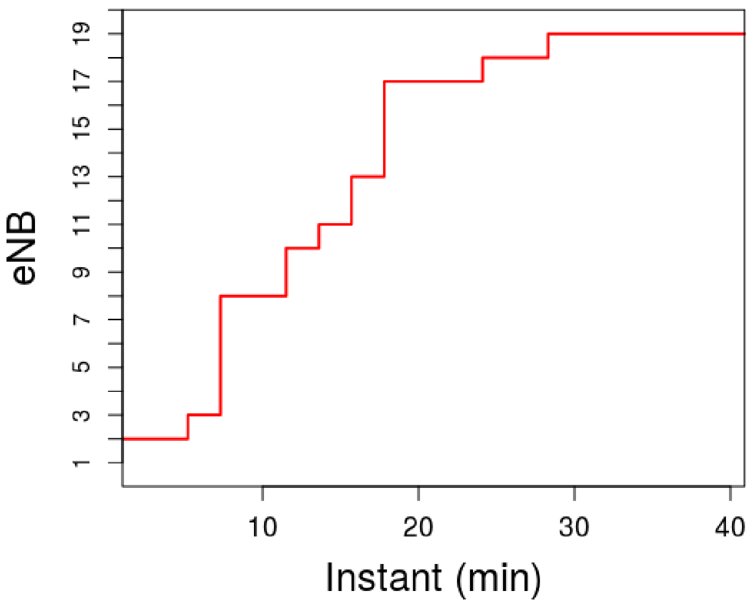


Figure 4.2: eNB in which the UE was allocated over the travel time.

Bandwidth management

To evaluate the model (COMP) bandwidth management, we simulated a second mobility scenario with 20 UEs and 5 eNBs. The bandwidth restrictions for eNBs and UEs are respectively L_i = 20 Mbps and l_k = 3 Mbps (hypothetical values). The model does not have time as a direct parameter, but the position (which influences the value of d_{ki}) changes over time. In this test case, UE's position x is given by the function $x = t$, and t represents a time instant. As previously illustrated in the flowchart of Figure 3.3, the model is executed periodically, at an instant t. Figure 4.3 shows the base station that the model allocates a user (namely UE_1). From the moment it becomes very costly to remain connected to a base station (because of distance or eNB handover average), e. g., eNB_1, the system will transfer the user to another eNB, Fig 4.3. The transition from eNB_1 to eNB_2 is performed near position $x = 30$.

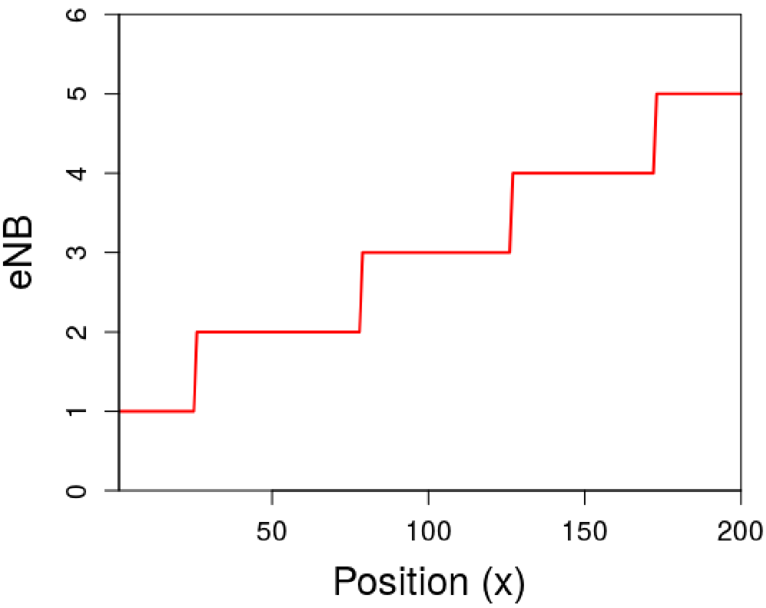


Figure 4.3: eNB_i in which UE_1 is allocated according to its position.

The process repeats as UE_1 moves towards other base stations. The remaining 19 UEs are not moving so we can easily see the behavior of the network. The graphs in Figure 4.4 and 4.5 show respectively the number of UEs allocated and the available bandwidth in each eNB, which changes as UE_1 is transferred between eNBs. We can see in Figure 4.4 that eNB_1 initially has 2 connected UEs. It is also known that UE_1 has its position determined by $x = t$. Therefore, at $t = 0$, UE_1 is at $x = 0$ and allocated in eNB_1. Observing the bandwidth availability of eNB_1 in Figure 4.5 at $t = 0$, for 2 connected users with minimum requirements of 3 Mbps, it is verifiable that the base station has only 14 Mbps, from a total L_1 = 20 Mbps.

We can verify the rearrangement of the network as UE_1 moves. As previously shown, the transfer of eNB_1 to eNB_2 occurs approximately at $x = 30$. We can see the change in the number of UEs in eNB_1 (from 2 to 1) and eNB_2 (from 3 to 4) in Figure 4.4 at $t \approx 30$. The gain or loss of

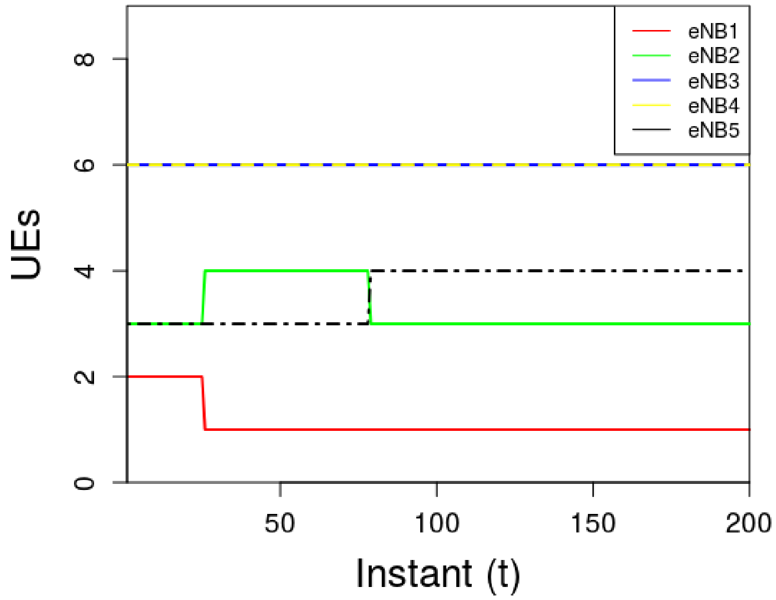


Figure 4.4: Number of connected UEs in each eNB over time.

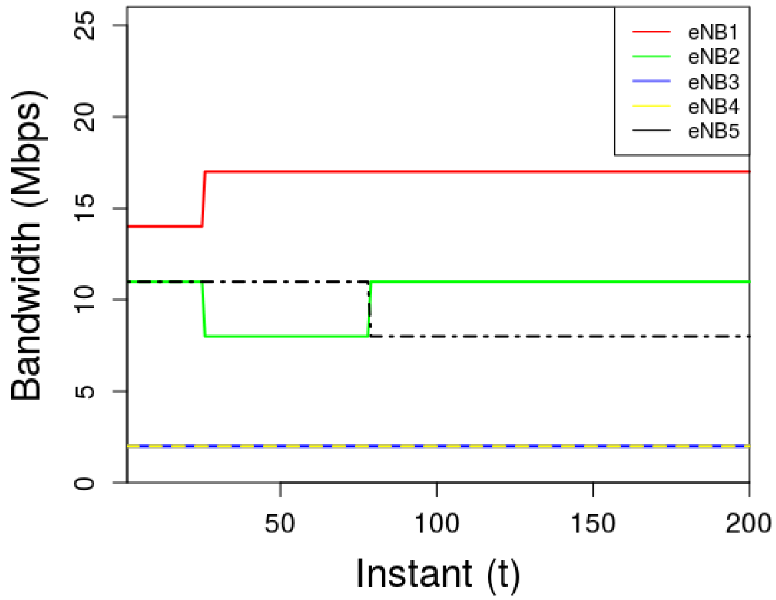


Figure 4.5: Available bandwidth in each eNB at instant t.

available bandwidth is observable in Figure 4.5 at the same instant t. Since eNB_4 and eNB_5 are always operating with 6 users, their bandwidth availability is low and insufficient to serve another UE. However, when approaching these base stations, UE_1 will certainly connect to them, as is observable in Figure 4.4. This behavior is possible because even if operating in full capacity, the network will rearrange UEs in different eNBs, ensuring a base station connection.

4.2 Heuristic performance

In this section, we present the results for the allocation model (3.19) considering RSRQ and the average handover frequency of each eNB. We will compare the results of the iterated local search, and a GRASP algorithm. To validate the results, we executed an exact model with the IBM ILOG CPLEX Optimization Studio V12.10, which provides the optimal solution through the branch-and-bound algorithm.

We created the problem's instances according to the standard GAP generation scheme (Chu and Beasley, 1997). They are divided per difficulty and number of variables, being **A** the easy instances, **B** moderate, and **C** presents the same difficulty as **B**, but with more variables. Table 4.2 presents the total number of UEs and eNBs in each instance.

Table 4.2: The number of UEs and eNBs according to the instance type.

Instance	Type	UEs Quantity	eNBs Quantity
1	**A/B**	100	5
2		100	10
3		100	20
4		200	5
5		200	10
6		200	20
1	**C**	1200	5
2		1200	10
3		1200	20
4		1500	5
5		1500	10
6		1500	20

Execution time

The algorithms were executed 30 times per instance on a computer with an Intel Core i7 @2.7 GHz x 4 processor, 16 GB of RAM, and the Ubuntu 18.04 LTS operating system. Table 4.3 summarizes the overall average execution time of the algorithms. The results presented in Table 4.4 contains the instances' optimal value provided by the branch and bound exact model, the average execution time of each solution, the best value achieved by the heuristic methods, the number of times that the best value was found $\mathbf{N}_{bests}$, and the average solution result.

The worst performing was the GRASP algorithm with a 3.38 s overall average execution time. The ILS heuristic achieves the best performance, with an 11.09 ms overall average execution time.

If we compare the average execution time of the exact model, 61.43 ms, with the ILS algorithm, there is a reduction in the average execution time by approximately 82%. Table 4.4 also

Table 4.3: Overall execution time results.

	Exact MD	**GRASP**	**ILS**
Overall Average Time (ms)	61.43	3381.76	11.09

shows that the ILS algorithm found the optimal solution for every instance. Column N_best indicates the number of times the instance evaluation and the algorithm found the optimal solution.

The GRASP solution presented the worst results, with an overall longer execution time when compared to the ILS and exact model solution. Also, it barely found the instances' optimal results. Briefly discussing The GRASP algorithm, its general structure contains a greedy function that starts by shuffling a list of users and base stations. We look for the base station that offers the minimum allocation cost for each user, generating a solution.

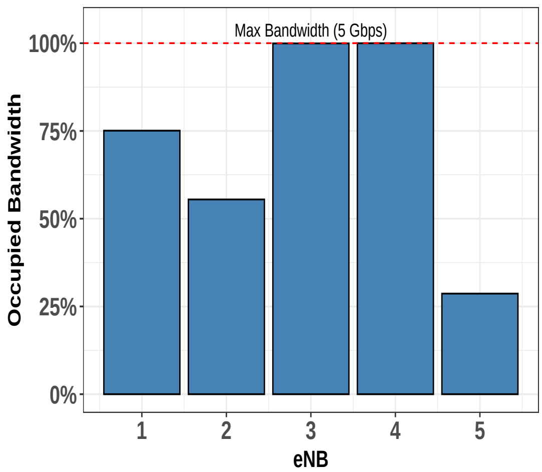
We apply the local search algorithm to improve the results, store it to a temporary best variable, and start the process again, looping until the algorithm reaches its time limit. The two key aspects that impact the algorithm's performance are the repeated generation of solutions per iteration, and the susceptibility to local minimum, since there is no diversification of perturbation mechanisms.

Table 4.4: The results for the exact, GRASP, and the ILS model.

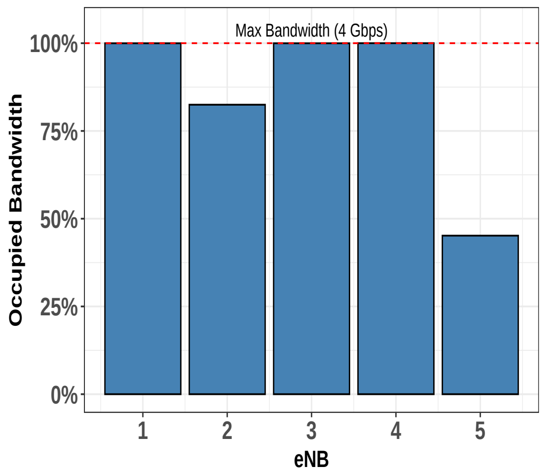
Instance	**Type**	**Exact MD**		**GRASP**				**ILS**			
		Optimal	**AVG Time (ms)**	**Best**	**Average**	**AVG Time (ms)**	**N$_{bests}$**	**Best**	**Average**	**AVG Time (ms)**	**N$_{bests}$**
1	A	101.77	5.80	101.77	101.80	3.04	13/30	101.77	-	0.18	30/30
2		94.82	9.65	94.87	95.02	4.14	0	94.82	-	0.21	30/30
3		89.17	16.34	89.21	89.32	4.30	0	89.17	-	0.24	30/30
4		209.67	8.36	209.67	209.87	31.48	5/30	209.67	-	0.72	30/30
5		179.97	17.46	180.04	180.18	33.00	0	179.97	-	0.67	30/30
6		169.15	31.09	169.24	169.39	36.52	0	169.15	-	1.02	30/30
1	B	102.64	5.35	102.64	102.71	2.44	15/30	102.64	-	0.19	30/30
2		91.96	8.68	91.96	92.06	3.07	4/30	91.96	-	0.22	30/30
3		88.38	19.37	88.40	88.54	4.15	0	88.38	-	0.26	30/30
4		235.75	7.91	235.75	235.77	26.90	23/30	235.75	-	0.88	30/30
5		197.79	19.53	197.87	198.03	32.23	0	197.79	-	0.77	30/30
6		158.65	34.29	158.74	158.86	35.20	0	158.65	-	1.00	30/30
1	C	1221.85	47.87	1221.85	1221.88	5880.62	22	1221.85	-	23.57	30/30
2		1057.96	109.02	1057.96	1058.10	7320.85	4	1057.96	-	25.15	30/30
3		1020.05	237.66	1020.22	1020.41	7638.71	0	1020.05	-	26.88	30/30
4		1636.13	76.29	1636.13	-	11998.06	30/30	1636.13	-	37.10	30/30
5		1397.8	143.22	1397.84	1398.15	13146.96	0	1397.8	-	38.28	30/30
6		1248.36	307.77	1248.59	1248.78	14670.06	0	1248.36	-	42.25	30/30

Bandwidth distribution evaluation

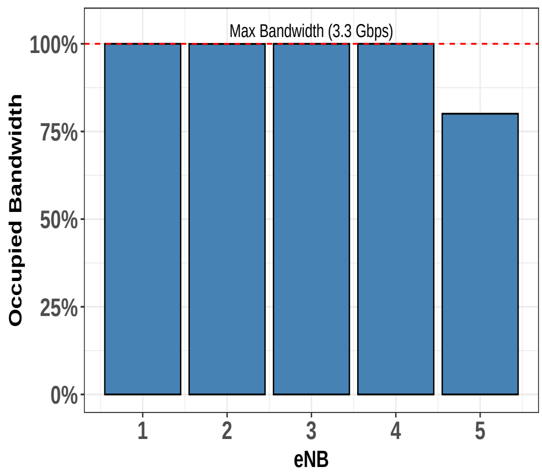
Evaluating new bandwidth management results, Figure 4.6 shows that as long as there is enough network bandwidth 4.6(a), we can freely allocate the users to the eNB which provides the best service. The network occupation rates are somewhat irregular. By decreasing the network available bandwidth per eNB while maintaining the same number of users and network requirements, we can see the occupation levels tend to have a more uniform distribution, which is more explicit in Figure 4.6(c). Under bandwidth restrained scenarios, the model will prioritize serving all the network users as best as possible at the sacrifice of service quality since we will allocate the users to secondary or tertiary transmission sources.



(a) 5 Gbps bandwidth maximum limit.



(b) 4 Gbps bandwidth maximum limit.



(c) 3.3 Gbps bandwidth maximum limit.

Figure 4.6: Network occupation levels for different max capacity limits per eNB.

4.3 Simulation of the proposed model

User allocation

In this section, we evaluate the user allocation provided by the ILS metaheuristic. The simulation process is similar to the flowchart presented in Figure 3.3, but this time we do not have the coverage area creation and data center communication cost step. The evaluation of the allocation process considers the region of São Paulo city, Brazil, shown in Figure 4.7. This region can be obtained through the OpenStreetMap (OSMF, 2021) website by informing the coordinates: $latitude_{min}$ = -23.558, $latitude_{max}$ = -23.5426, $longitude_{min}$ = -46.642, $longitude_{max}$ = -46.6249. We can use this region's perimeter data to extract a set of real eNBs positions from the Telebrasil (2021) base station dataset. The simulation region has a 1947.65 m x 1878.95 m area and contains 26 eNBs from a local phone carrier. For our experiment, we will consider a 300 meters transmission range, and the base stations are labeled from 0 to 25. We use the SUMO (Simulation of Urban MObility) simulator to generate the route shown in Figure 4.8, which has 432 positions. The UE starts at the bottom left and proceeds to the final destination at the top right.

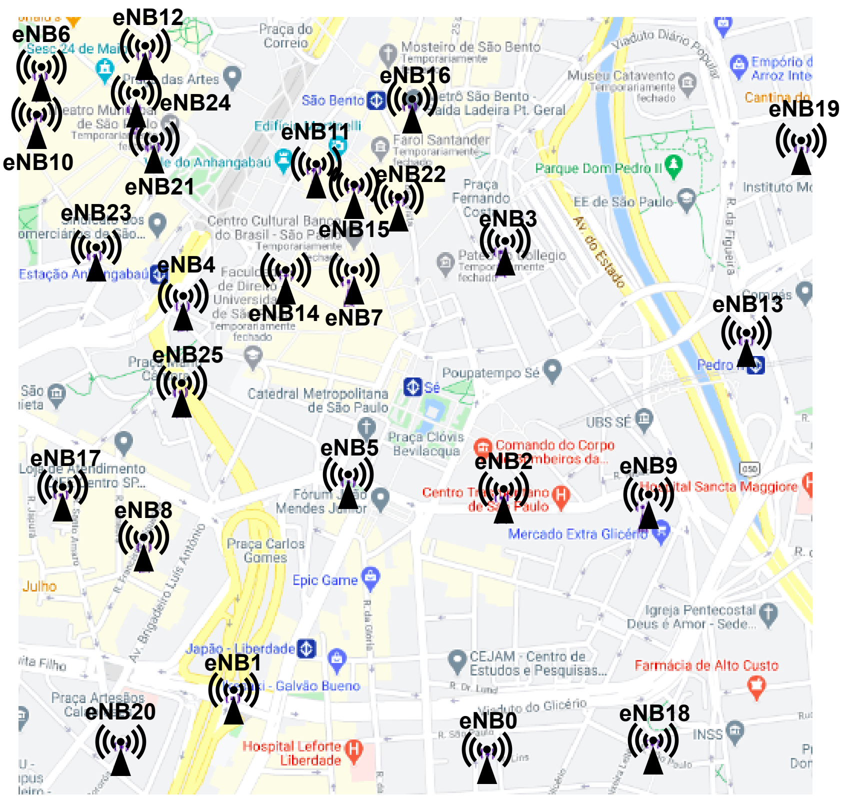


Figure 4.7: Experiment's eNBs positions at the city of São Paulo, Brazil.

Figure 4.8: UE simulated route generate with SUMO.

Figure 4.9 shows the eNBs in which we allocate the user throughout the route (highlighted in green). We can see that the allocations are very similar, differing for a moment around the time-step 200. Our model kept the UE connected for a longer time in eNB$_5$, but the number of handovers had not changed as the final result. Each model performed 7 handovers if we discard the initial allocation at eNB$_{20}$ (the route starts at the bottom left).

The mobility simulation process reads one position per iteration until the final destination (top right). The models receive the positions to compute the best candidate eNB. Both Ahmadi et al. (2020) and our solution are affected by the RSRQ. The farther away, the worst the RSRQ value. The eNB's average handover frequency h_i and the RSRQ value Θ_{ki} (which is a function of the distance that we compute with the UE's current position) are enough for our allocation model to compute the best candidate eNB, as established in Equation (3.19).

On the other hand, Ahmadi et al. (2020) proposal requires all the possible routes connecting a user's current position and final destination to perform the allocation. For simplification purposes, this simulation only considered the route of Figure 4.8 as the possible path, which is a best-case scenario where we can calculate the ideal point as the route's middle point, which changes every time the current position is updated.

Figures 4.10(a) and 4.10(b) show respectively the allocation sequence for Ahmadi et al. (2020) and our proposal, with the period of each allocation. As previously stated, our solution maintained the user in eNB$_5$ for a longer period. Each model performed 7 handovers, starting

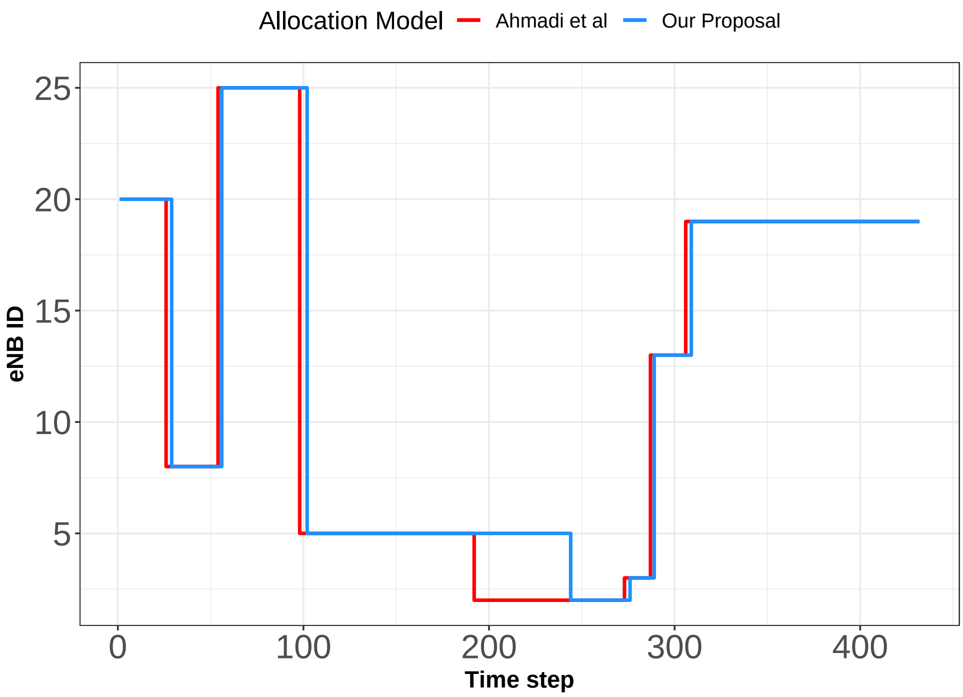
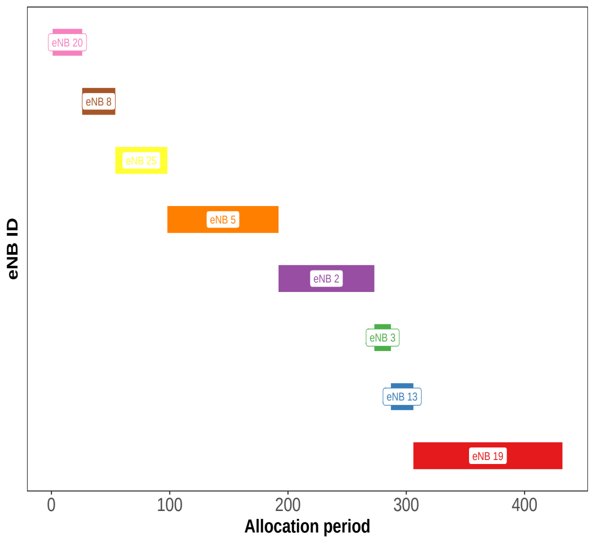
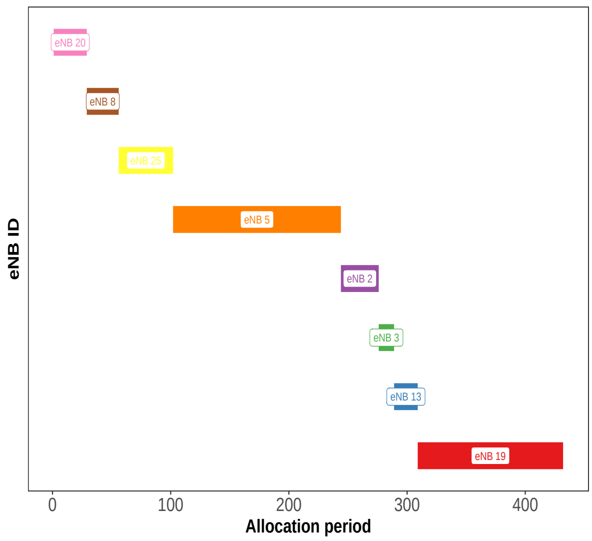


Figure 4.9: User allocation through the route.

from eNB_{20}, proceeding to eNB_8 and continuing until eNB_{19}.



(a) Ahmadi's model



(b) Our proposal.

Figure 4.10: UE's allocation eNBs for each model.

On the evaluation of other metrics besides the allocation and handover behavior, Figure 4.11 shows the user's RSRQ status through the route. The closer to zero, the better. It is in our interest to always choose the best RSRQ eNB to maintain the users' connection quality. The models' results are mostly overlapped through the route, with short periods when our proposal is better. We can now understand why we kept the user allocated in eNB_5 for a more extended period around the 200 time-step mark since this eNB provides a better RSRQ.

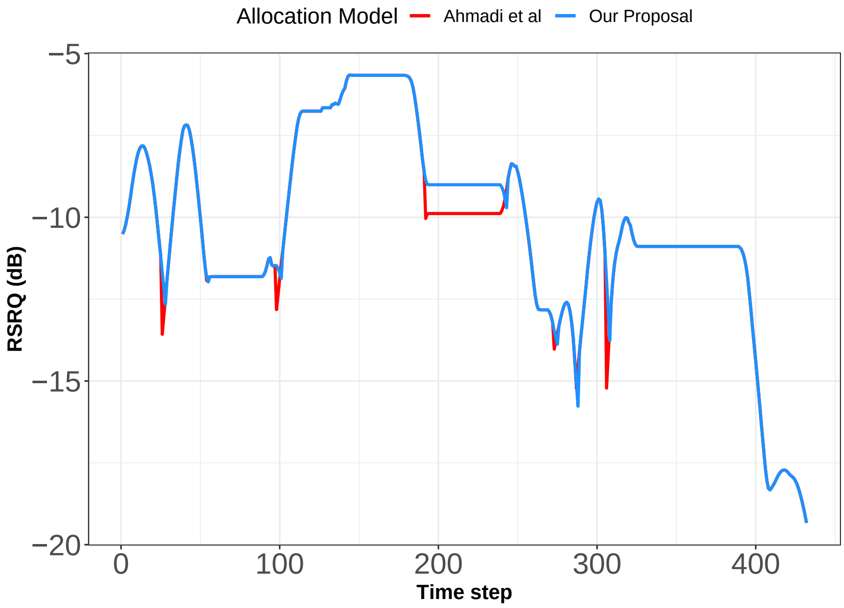


Figure 4.11: User's RSRQ connection value through the route.

According to Figure 4.11, achieved better RSRQ results for short periods, but to investigate further and reduce the influence of the eNBs position, we generated 100 instances with different random eNBs placement. We consider the São Paulo map region of Figure 4.7, 26 eNBs, and the same user's travel route 4.8. For each instance, we save travel's RSRQ time series and calculate its average value.

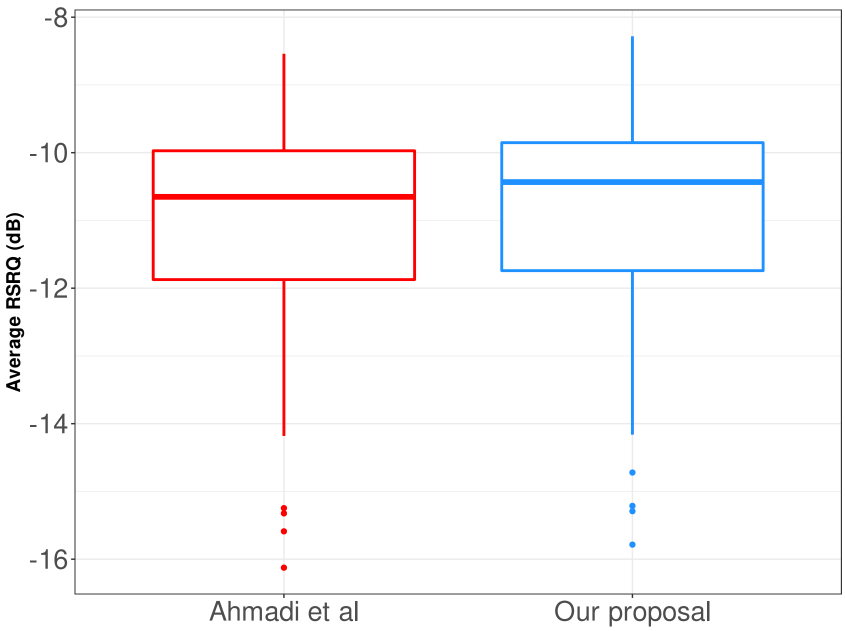


Figure 4.12: Boxplot of the average RSRQ value for 100 random instances of eNB placement.

Figure 4.12 presents the boxplot of the RSRQ averages for each placement instance. The overall RSRQ average for Ahmadi's and our proposal are -11.03 dB, and -10.87 dB, which gives a 1.45% gain to our allocation model. The boxplot shows that the models produce very similar solutions, and the two-sided Kolmogorov-Smirnov test considering a 5% significance provides p-value = 0.3667 (> 0.05), which reinforces that the models are statistically equal. The advantage of our solution is that our allocation considers the network's bandwidth limits, and also, we do not need to know or predict the users' future routes. Instead, we require only the current position.

Ahmadi et al. (2020) rely on third-party applications to compute, for every network user, all possible routes to reach the final destination. These routes are the inputs to calculate an ideal point that aims to cover most paths, and the eNBs closer to the ideal point receive higher scores than distant ones.

Realistically, this approach may face severe time response overhead due to third-party applications network delay. Furthermore, finding all the possible paths between two points is a $\mathcal{NP}$-hard problem (Chatterjee and Banerjee, 2014). Performing this operation for each user is an expensive task.

This chapter presented the results of the allocation model. In Section 4.1, we evaluated the average execution time, user allocation, and bandwidth management for the distance-based model solved with the branch and bound algorithm. In Section 4.2, we discussed the results of our heuristic approach considering the RSRQ parameter. Similar to the previous evaluation, we observed the heuristic approach average execution time, bandwidth management according to base station's bandwidth occupation, and allocation results were discussed in Section 4.3.

5

FINAL CONSIDERATIONS

5.1 Conclusion

The transition to the 5G generation brings new possibilities for improving application and models. In this work, we presented a mathematical model to improve allocation and handover processes considering users (UEs) and base stations (eNBs) of mobile networks. Our first proposal considered the shortest distance between base stations and users, the handover average, and the bandwidth requirements. Since distance does not consider wireless communication factors such as noise and environment interference, we replaced the distance parameter with the RSRQ (Reference Signal Received Quality) indicator, which measures communication quality between users and base stations.

As the main mobility simulation scenarios, we considered a 13.7 Km route at Maceió City, Brazil. A user provides its current location at the route with GPS collected data. In the second mobility scenario, we use a map region of the city of São Paulo, Brazil. The map region has the real position of 26 eNBs from a local phone carrier, and we simulate the user routes with the SUMO simulator.

We implemented the allocation models with a heuristic method and the CPLEX optimization solver, which solves linear integer problems with the branch and bound algorithm, an exact solution. In resource-constrained scenarios, exact algorithms might exceed the limits of available resources. Heuristic approaches are better for these situations, and at the sacrifice of optimal values for good enough solutions, we can obtain better resource management and faster-solving methods.

On average, the iterated local search obtained an execution time reduction of approximately 82% compared to the branch and bound exact algorithm. The GRASP-based algorithm showed the worst results due to its recurrent solution construction and lack of mechanisms to avoid local minimum. Regarding the RSRQ indicator, the solution reached a 1.45% average gain, and the number of performed handovers was maintained, compared to a similar literature model. Despite the modest improvement, which makes our proposal statistically equivalent to the literature model, we offer the advantage of not predicting the users' possible and future routes. Only the current position is required. Furthermore, our solution also considers base stations' bandwidth capacity, controlling the allocation and network occupation limits.

We conclude by stating the following contributions: i) A mathematical model to allocate users in base stations of mobile wireless networks; ii) Proposition and evaluation of a heuristic method as an alternative to exact solving approach; iii) Evaluation of allocation on different mobility simulations, one of which considers the real position of base stations from a local mobile carrier; iv) A evaluation and comparison of our model with existing literature work. Also, some of our work results have already been published (Ramos et al., 2019b,a).

5.2 Future work

Designing allocation models to wireless networks, we can consider characteristics such as prediction factors, user's traveling speed, and estimation of failure probability. Future work includes the study of new characteristics and their impacts on the mathematical formulation. We consider the development of a hybrid approach combining the advantages of ILS-VND with an exact method. Heuristic approaches can provide satisfactory results, but it is susceptible to non-optimal values. We can improve our proposal by generating initial solutions with heuristic methods and using the result as an input to exact methods. This approach would impact the average execution time of the exact method and provide an optimal solution. As for new mobility scenarios, we consider integration and real-time simulation with the SUMO mobility simulator.

REFERENCES

K. Ahmadi, S. Pourya Miralavy, and M. Ghassemian. Software-defined networking to improve handover in mobile edge networks. *International Journal of Communication Systems*, page e4510, jun 2020.

I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002.

Zoraze Al, Nicola Baldo, Josep Mangues-Bafalluy, and Lorenza Giupponi. Machine learning based handover management for improved QoE in LTE. *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, jul 2016.

Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys & Tutorials*, 17(4):2347–2376, jun 2015.

David L. Applegate, Robert E. Bixby, Vašek Chvatál, and William J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2006. ISBN 9780691129938.

John Baxter. Local optima avoidance in depot location. *Journal of the Operational Research Society*, 32(9):815–819, 1981.

Federico Boccardi, Jeffrey Andrews, Hisham Elshaer, Mischa Dohler, Stefan Parkvall, Petar Popovski, and Sarabjot Singh. Why to Decouple the Uplink and Downlink in Cellular Networks and How To Do It. *IEEE Communications Magazine*, 54(3):110–117, mar 2016.

Yulian Cao, Han Zhang, Wenfeng Li, Mengchu Zhou, Yu Zhang, and Wanpracha Art Chaovalitwongse. Comprehensive learning particle swarm optimization algorithm with local search for multimodal functions. *IEEE Transactions on Evolutionary Computation*, 23(4): 718–731, 2018.

Dirk G Cattrysse and Luk N Van Wassenhove. A survey of algorithms for the generalized assignment problem. *European journal of operational research*, 60(3):260–272, 1992.

Sankhadeep Chatterjee and Debarshi Banerjee. A novel boolean expression based algorithm to find all possible simple paths between two nodes of a graph. *International Journal of Advanced Research in Computer Science*, 5(7), 2014.

Min Chen, Shiwen Mao, and Yunhao Liu. Big Data: A Survey. *Mobile Networks & Applications*, 19(2):171–209, APR 2014.

Paul C Chu and John E Beasley. A genetic algorithm for the generalised assignment problem. *Computers & Operations Research*, 24(1):17–23, 1997.

CISCO. Cisco Annual Internet Report (2018–2023). https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf, mar 2020. White paper, Retrieved June 02, 2020.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, third edition, 2009.

Luis M. Correia, Dietrich Zeller, Oliver Blume, Dieter Ferling, Ylva Jading, István Gódor, Gunther Auer, and Liesbet Van Der Perre. Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks. *IEEE Communications Magazine*, 48(11):66–72, nov 2010.

Aleksandar Damnjanovic, Juan Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, Madhavan Vajapeyam, Taesang Yoo, Osok Song, and Durga Mallad. A Survey on 3GPP Heterogeneous Networks. *IEEE Wireless Communications*, 18(3):10–21, jun 2011.

George B Dantzig. Origins of the simplex method. In *A history of scientific computing*, pages 141–151. ACM, 1990.

Xiaoyu Duan and Xianbin Wang. Authentication handover and privacy protection in 5G hetnets using software-defined networking. *IEEE Communications Magazine*, 53(4):28–35, apr 2015.

M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

GSMA Intelligence. Understanding 5G: Perspectives on future technological advancements in mobile. https://www.gsma.com/futurenetworks/wp-content/uploads/2015/01/Understanding-5G-Perspectives-on-future-technological-advancements-in-mobile.pdf, dec 2014. White paper, Retrieved June 04, 2020.

Pierre Hansen, Nenad Mladenović, Jack Brimberg, and José A. Moreno Pérez. Variable neighborhood search. In *Handbook of Metaheuristics*, pages 57–97. Springer International Publishing, sep 2018.

IBM. IBM ILOG CPLEX Optimization Studio Getting Started with CPLEX. https://www.ibm.com/support/knowledgecenter/SSSA5P_12.8.0/ilog.odms.studio.help/pdf/gscplex.pdf, apr 2017. Online, Retrieved June 04, 2020.

Zhenyue Jia, Jianqiao Yu, Xiaolin Ai, Xuan Xu, and Di Yang. Cooperative multiple task assignment problem with stochastic velocities and time windows for heterogeneous unmanned aerial vehicles using a genetic algorithm. *Aerospace Science and Technology*, 76:112–125, may 2018.

James E Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.

Victor Klee and George J Minty. How good is the simplex algorithm. *Inequalities*, 3(3):159–175, 1972.

Peter J. Kolesar. A branch and bound algorithm for the knapsack problem. *Management science*, 13(9):723–735, may 1967.

Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements*, 5(3&4), 2012.

Slawomir Kuklinski, Yuhong Li, and Khoa Truong Dinh. Handover management in SDN-based mobile networks. *2014 IEEE Globecom Workshops, GC Wkshps 2014*, pages 194–200, mar 2015.

Ailsa H Land and Alison G Doig. An automatic method for solving discrete programming problems. In *50 Years of Integer Programming 1958-2008*, pages 105–132. Springer, 2010.

Eugene L Lawler and David E Wood. Branch-and-bound methods: A survey. *Operations research*, 14(4):699–719, 1966.

Jiseong Lee and Younghwan Yoo. Handover cell selection using user mobility information in a 5G SDN-based network. *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 697–702, jul 2017.

Helena Ramalhinho Lourenço, Olivier C Martin, and Thomas Stützle. Iterated local search: Framework and applications. In *Handbook of metaheuristics*, pages 129–168. Springer, 2019.

R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, apr 2004.

Olivier Martin, Steve W Otto, and Edward W Felten. *Large-step Markov chains for the traveling salesman problem*. Citeseer, 1991.

Olivier C Martin and Steve W Otto. Combining simulated annealing with local search heuristics. *Annals of operations research*, 63(1):57–75, 1996.

Peter Mell and Tim Grance. The NIST Definition of Cloud Computing. *Communications of The ACM*, 53(6):50, jun 2010.

Nenad Mladenović and Pierre Hansen. Variable neighborhood search. *Computers & operations research*, 24(11):1097–1100, 1997.

Saraju P. Mohanty, Uma Choppali, and Elias Kougianos. Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5 (3):60–10, jul 2016.

Bruno Astuto A. Nunes, Marc Mendonca, Xuan-Nam Nguyen, Katia Obraczka, and Thierry Turletti. A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. *IEEE Communications Surveys & Tutorials*, 16(3):1617–1634, feb 2014.

Temel Öncan. A survey of the generalized assignment problem and its applications. *INFOR: Information Systems and Operational Research*, 45(3):123–141, 2007.

OSMF. OpenStreetMap. https://www.openstreetmap.org/about, october 2021. Online, Retrieved October 04, 2021.

Manfred Padberg and Giovanni Rinaldi. Optimization of a 532-city symmetric traveling salesman problem by branch and cut. *Operations Research Letters*, 6(1):1–7, 1987.

Nisha Panwar, Shantanu Sharma, and Awadhesh Kumar Singh. A survey on 5G: The next generation of mobile communication. *Physical Communication*, 18:64 – 84, 2016. Special Issue on Radio Access Network Architectures and Resource Management for 5G.

Puca Huachi Vaz Penna, Anand Subramanian, and Luiz Satoru Ochi. An iterated local search heuristic for the heterogeneous fleet vehicle routing problem. *Journal of Heuristics volume*, 19(2013):201–232, sep 2011.

Jonathan Prados-Garzon, Oscar Adamuz-Hinojosa, Pablo Ameigeiras, Juan J Ramos-Munoz, Pilar Andres-Maldonado, and Juan M Lopez-Soler. Handover implementation in a 5g sdn-based mobile network architecture. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6. IEEE, 2016.

Li Qiang, Jie Li, and Changcheng Huang. A software-defined network based vertical handoff scheme for heterogeneous wireless networks. *2014 IEEE Global Communications Conference*, 2014.

Li Qiang, Jie Li, and Corinne Touati. A User Centered Multi-Objective Handoff Scheme for Hybrid 5G Environments. *IEEE Transactions on Emerging Topics in Computing*, 5(3): 380–390, apr 2016.

Geymerson S. Ramos, Rian Pinheiro, and André L. L. Aquino. Otimização de Processos e Mobilidade para Redes Móveis 5G com Redes Definidas por Software. In *LI Simpósio Brasileiro de Pesquisa Operacional*, volume 2, sep 2019a.

Geymerson S. Ramos, Rian G. S. Pinheiro, and André L. L. Aquino. Optimizing 5G Networks Processes With Software Defined Networks. *IEEE 8th International Conference on Cloud Networking (CloudNet)*, nov 2019b.

G Terry Ross and Richard M Soland. A branch and bound algorithm for the generalized assignment problem. *Mathematical programming*, 8(1):91–103, 1975.

Sartaj Sahni and Teofilo Gonzalez. P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565, 1976.

Harvey M Salkin and Cornelis A De Kluyver. The knapsack problem: a survey. *Naval Research Logistics Quarterly*, 22(1):127–144, 1975.

J. K Sharma. *Operations Research: Theory and Application*. LAXMI PUBLICATIONS; Sixth Edition, sixth edition, jan 2017.

David B Shmoys and Éva Tardos. An approximation algorithm for the generalized assignment problem. *Mathematical programming*, 62(1):461–474, 1993.

Roger W. Sinnott. Virtues of the Haversine. *Sky Telescope*, 68:159, 1984.

Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

Tassin Srivarapongse and Phajongjit Pijitbanjong. Solving a special case of the generalized assignment problem using the modified differential evolution algorithms: a case study in sugarcane harvesting. *Journal of Open Innovation: Technology, Market, and Complexity*, 5 (1):5, 2019.

Tarik Taleb, Miloud Bagaa, and Adlen Ksentini. User mobility-aware Virtual Network Function placement for Virtual 5G Network Infrastructure. In *2015 IEEE International Conference on Communications (ICC)*, 2015.

Telebrasil. Mapa de ERBs Brasil (antenas). `http://www.telecocare.com.br/telebrasil/erbs/`, 2021. Retrieved October 04, 2021.

Ms. Lopa J. Vora. Evolution of Mobile Generation Technology: 1G to 5G and Review of Upcoming Wireless Technology 5G. *International Journal of Modern Trends in Engineering and Research*, 2(10):281–290, oct 2015.

Jun Wu, Zhifeng Zhang, Yu Hong, and Yonggang Wen. Cloud Radio Access Network (C-RAN): A Primer. *IEEE Network*, 29(1):35–41, jan 2015.

Xiaohuan Yan, Y. Ahmet Şekercioğlu, and Sathya Narayanan. A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks. *Computer Networks*, 54(11):1848–1863, aug 2010.

Haijun Zhang, Na Liu, Xiaoli Chu, Keping Long, Abdol-Hamid Aghvami, and Victor C. M. Leung. Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges. *IEEE Communications Magazine*, 55(8):138–145, aug 2017.