POLITECNICO DI TORINO

Master's Degree in ICT FOR SMART SOCIETIES

Master's Degree Thesis

Sustainability in Mobile Networks: Analyzing the Feasibility and Benefits of Infrastructure Sharing

Supervisors

Prof. Michela MEO

Prof. Daniela RENGA

Candidate

Maoquan NI

March 2024

Summary

Introduction

The advent of 5G technology brings higher data rates, stricter latency requirements, and greater connectivity, presenting significant challenges for the deployment of future communication networks. As traffic volumes rise, concerns over the energy consumption and efficiency of mobile access network infrastructures intensify. These infrastructures are designed to meet peak demands but often remain underutilized most of the time, leading to substantial energy waste and financial costs. This thesis introduces a novel strategy known as network sharing (NS), where two or more mobile network operators (MNOs) agree to collaboratively share their radio access network (RAN) infrastructure. This approach aims to reduce energy usage by minimizing the active periods of infrastructure during off-peak traffic demand, thereby maximizing energy efficiency without significantly compromising the quality of service.

The thesis delves into network sharing dynamics within telecommunication infrastructure management, focusing on cooperative operations among base stations for optimal resource utilization. By deactivating specific base stations and leveraging the capacity of neighboring units to handle the mobile traffic offloaded from the switched-off network nodes, the study seeks to achieve energy efficiency and reduce operational costs. A case study involving base stations owned by two different MNOs illustrates the asymmetric sharing dynamics, highlighting the impact of variable capacities on the feasibility and efficiency of network sharing arrangements.

Methodology

This study adopts a data-driven methodology to evaluate network sharing (NS) potential and efficiency, aiming for energy conservation and financial savings across diverse urban and rural settings. Utilizing the real normalized mobile traffic profiles from the NetMob dataset provided by a French mobile operator, this research encompasses a broad range of digital activities over 77 days. The data, capturing

traffic loads from more than 68 mobile services, are organized by geographical tiles and documented for each 15-minute interval of the day. This extensive dataset facilitates a nuanced analysis of traffic distribution patterns critical for network sharing feasibility and optimization.

The methodological approach involves several key stages, beginning with data aggregation and preprocessing of a month's worth of traffic load data. Selection criteria for mobile network operators and geographic areas focus on stations operated by major providers across demographically diverse regions. Traffic load time series, that are available on a per-tile basis are allocated to the nearest base station using the Euclidean distance method, enabling the construction of individual base station aggregated traffic traces. The feasibility of network sharing is then evaluated by analyzing base station capacity and a preset threshold for sharing initiation (this threshold defines the maximum saturation level of the capacity of the BS hosting the offloaded traffic), determining the time slots that could implement NS while maintaining service quality.

A predefined power consumption model calculates potential energy and economic savings, expanding from dual to triple base station configurations to further enhance energy savings . Additionally, the frequency of base station on/off switching operations is critically examined for the feasibility and practicality of NS strategies. Performance metrics and predictive analysis using machine learning techniques to forecast future traffic patterns allow for proactive NS arrangements.

The dataset's depth allows for exploring NS policies based on actual data about base station locations and traffic loads. Through a careful process, traffic data from each specified block are aligned to its corresponding base station, considering both the precise locations of the base stations and the regional traffic load data. The study's focus on modeling raw traffic traces, determining NS thresholds, defining base station capacity, and delineating area-specific strategies highlights the comprehensive nature of this methodology. It integrates the principles of the power consumption dynamics of base stations, particularly focusing on Remote Radio Heads (RRHs), where a near-linear relationship between RF output power and power consumption is established.

This methodology represents a significant step toward realizing network sharing within base station infrastructure, providing a foundation for evaluating the energy savings attributable to NS. By synthesizing traffic prediction models like LSTMs, XGBoost, and ARIMA, the study navigates the intricacies of traffic load forecasting, leveraging these insights to optimize network sharing arrangements and contribute significantly to telecommunications infrastructure's energy efficiency and sustainability.

Results and Discussion

- This thesis has made significant contributions to the understanding and strategic planning and management of mobile network infrastructure, with a particular emphasis on the analysis of base station traffic traces and their geographical distribution, which will influence the NS potential. It aimed to improve network planning and infrastructure sharing in scenarios with both double and triple base station pairs owned by different operators.
  - In the context of double base station (BS) Network Sharing, the findings indicate that NS is feasible for 60-80% of the time, yielding power savings in the range of 25% to 35%. A derived linear regression model, represented by the equation power saving = $0.48 \cdot$ NS percentage - 0.04, provides a quantitative basis for estimating the impact of NS utilization on energy conservation.
  - The triple BSs NS with two saturation thresholds improves the NS percentage to 90% and increases power savings by 15% compared to the same BS with double NS. The resultant substantial energy savings have the potential to significantly reduce electrical costs, contributing towards broader energy conservation efforts.
- The frequency of switch operations is thought as a critical factor in evaluating the practicality of implementing a NS strategy in real-world scenarios. This section's research will focus on developing strategies to reduce the frequency of these operations, hence preventing fast degradation of network nodes.
  - The fixed period strategies that first minimize rapid switch operations in network sharing (NS), introduces "forbidden fast changes" to extend NS duration to at least 30 minutes, effectively eliminating short-term switching patterns. This approach significantly reduces the frequency of state changes, maintaining the efficacy of NS utilization. Further enhancements included setting limits on switch operations within each 6 hours period (each day is divided into four six-hour periods: 00:00-06:00, 06:00-12:00, 12:00-18:00, and 18:00-24:00), ensuring that after a certain number of switches, the NS status is uniformly set to inactive for the remainder of the period. These measures not only simplify NS operations but also maintain the feasibility of network sharing among mobile network operators (MNOs).
  - Another method has been discovered, which is based on the concept of a sliding window that moves across NS state time intervals. If the frequency of status changes during that window exceeds a predefined threshold, all statuses until the sliding window ends are changed to a state that

terminates NS. When comparing the energy trade-off parameter, the fixed period strategy appears to be the better option, maintaining the majority of the energy savings while limiting switch operations. Furthermore, fixed-period strategies provide a consistent operational schedule, making it easier to adjust to real-time data.

- Furthermore, the investigation has utilized time series machine learning techniques, including LSTM, XGBoost, and ARIMA, to forecast traffic loads and subsequent network sharing utilization.
  - Among these methods, ARIMA showed marginally superior performance in predicting peak traffic loads.
  - The predictive accuracy for NS utilization was approximately 85%, an analysis was conducted to quantify the traffic load at points of false negatives and determine their excess over the capacity of the base station that is set in our NS strategy, which is traditionally set at 80% of the base station capacity. If the traffic load during false negative occurrences falls within the 80%-100% range, it suggests that while traffic slightly exceeds the saturation threshold, it does not exceed the BS's capacity, making NS viable without risk of overload, these are acceptable prediction results, and after adjusting 'fake false positives', the accuracy could potentially increase to around 93%.
  - These findings emphasize the accuracy of ML methods in forecasting NS applicability and network management, which ensures consistent delivery of high-quality services to users and at the same time improves the operational feasibility of NS strategies.

The findings from this research offer valuable insights into the optimization of network infrastructure and emphasize the viability of infrastructure sharing as a sustainable and efficient approach for the future of mobile networks. This thesis not only addresses key challenges in mobile network management but also lays a foundation for more sustainable and efficient telecommunications infrastructures. The strategies and methodologies proposed herein promise to revolutionize network sharing practices.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Michela Meo and Daniela Renga, for their solid support and guidance throughout my research. Their expertise and insights have been invaluable, and I am deeply grateful for their patience and encouragement. During this thesis research, we collaborated on two papers: one for the Netmob data challenge and another for the ICC conference 2024. These processes have greatly benefited me, and the successes will serve as motivation to continue working hard in the future.

I am also thankful to my colleague Song Tailai for constructive discussions, and shared experiences. Working alongside him has been both inspiring and motivating.

I'd like to express my gratitude to my family for their constant support and encouragement throughout my academic journey. Their faith in me has been a constant source of strength and encouragement, and this accomplishment is as much theirs as mine.

To my girlfriend, Wei Zhixuan, I am extremely thankful for her understanding, patience, and constant encouragement. Her encouragement and belief in me have been a guiding light during difficult times, and I am incredibly fortunate to have her in my life.

Finally, I'd like to thank TWICE for using their songs to give me strength during so many exhausting and difficult moments.

by Ni maoquan

Table of Contents

List of Tables

List of Figures

Acronyms

MNOs
Mobile network operators

BSs
base stations

NS
network sharing

QOS
quality of service

ML
machine learning

LSTM
long short-term memory

ARIMA
autoregressive integrated moving average

XGBoost
eXtreme Gradient Boosting

SC
state changes

SE
switch effiency

Chapter 1

Introduction

1.1 Overview

The advent of 5G technology marks a significant progress in mobile communication, offering unprecedented access to information and data transfer speeds. This advancement, however, comes with its own challenges. The deployment of 5G requires a greater number and density of base stations (BSs) than its predecessor, 4G. It is predicted that by 2025, there will be about 13.1 million BSs in the world, and the BS energy consumption will reach 200 billion kWh. [1] Additionally, the power consumption per base station increases by nearly 70% over a base station deploying a mix of 2G, 3G, and 4G radios, all of which leads to increased resource consumption and concerns about the sustainability of communication systems. Hence, even a 10 percent reduction in cellular network power consumption represents a savings of 20 billion kWh, which would save about US$5.8 billion—a substantial amount. [2] Given that base station infrastructures are designed to meet peak demands and often remain underutilized during 75%-90% of the time, this underutilized period presents an ideal opportunity to implement green network strategies.

Building on the concept of base station sleep mode for single Mobile Network Operator (MNO) base stations, this thesis examines the feasibility and advantages of an innovative approach to base station infrastructure sharing among different MNOs. The traditional model of network operation, where each MNO independently manages its Radio Access Network (RAN) infrastructure, is not optimized for energy efficiency. The motivation of this thesis is to propose a novel strategy for network sharing (NS) among MNOs to address the challenges of energy consumption and inefficiency in base station infrastructure. By collaborating to share RAN infrastructures that are geographically proximate, and effectively cover the same regions, it is possible to maintain Quality of Service (QoS) while

dynamically offloading traffic from one base station to another during off-peak hours. A threshold for infrastructure capacity is set to determine the feasibility of implementing the sharing strategy at any given time. After offloading, one base station can be temporarily deactivated, allowing the traffic originally managed by two operators to be supported by a single operator's base station. While satisfying users' QoS requirements, this strategy aims to reduce the operational time of underutilized infrastructure, particularly during off-peak hours, thereby enhancing energy efficiency and curtailing unnecessary financial expenditures.

Chapter 2

Background and related work

5G networks are expected to provide excellent quality of service (QoS) to a large number of devices, enable flawless functionality during user mobility, and improve energy efficiency when compared with previous communication technologies [3] [4]. Currently, 5G technology is in its early commercialization stages [4]. However, we continue to experience relevant challenges in the actual realization of the 5G era, which is defined by the widespread penetration of extremely demanding communication services based on massive Machine Type Communication, Ultra Reliable Low Latency Communications enhanced Mobile Broadband service categories, and the need for edge caching and computing to support smart mobility [5]. The advanced BS sleeping mode is a key method for reducing energy consumption. Some papers have already researched the efficiency and feasibility of the BS sleeping model, demonstrating that it could be a solution for a green and sustainable network. [6] [7] Another promising solution to the energy consumption problem is energy harvesting (EH), which involves collecting energy from renewable sources such as solar panels and wind turbines. Solar and wind energy are the most encouraging, environmentally friendly, and rapidly expanding clean and renewable energy sources for addressing such energy issues and saving the environment. [8][9]

In this context, sharing network resources among different MOs may play an important role in improving the energy efficiency and resilience of future mobile networks [5] [10] [11]. The goal of our research activity is to investigate the potential benefits in terms of sustainability and resilience derived from sharing Base Station (BS) usage among different MOs[12], which is similar to the idea of [13] in a macro-cell system with a number of small cells, and to propose a Distributed Base Station On/Off Control Mechanism (D-OCM) for reducing the network's power consumption. Using D-OCM, an inactive small base station (SBS) can be activated

by either an active SBS or the macro base station (MBS), while an active SBS can be deactivated by the active SBS itself, both in a distributed manner.

Before the research, some background of base station distribution is researched. Our dataset is a geographical distribution traffic load that represents for each small block in the metropolitan area. To analyze base station infrastructure sharing, we need to distribute the traffic load to base stations to build a traffic trace. The spatial distribution of base stations (BSs) and traffic demands is critical for effective network planning and infrastructure sharing, both of which are important components of green cellular networking. The distribution pattern of BS locations and neighbor distances was analyzed using massive real data of BS locations in a representative French metropolis. Using Delaunay triangulation and fine classification, the BS distribution is shown to be non-Poisson, with repulsive interaction between BS. The Strauss-Hardcore point process is the closest model to the actual BS distribution. [14] Another study discovered that the spatial distribution of BSs exhibits not only high nonuniformity across a region, but also diverse patterns in different regions, implying that the widely used homogeneous Poisson point process can only approximate BS patterns in a small area. As a result, the inhomogeneous PPP (IPPP), specifically the Cox point process with spatially varying intensity, is used to model the BS distribution at all spatial scales.[15]

And, in order to prepare operational adjustments in advance rather than in real time, some machine learning method should be used to predict traffic traces. The network traffic load prediction on base stations was modelled as a time series forecasting problem. Recurrent neural networks, such as LSTMs and GRUs, typically offer better performance and Clustering base stations based on their behavior before using time series forecasting methods to predict their loads can significantly improve load forecasting accuracy when there are enough records in each cluster.[16] Another research Using the SARIMA model, the most accurate prediction techniques for time series traffic load are presented. [17] During my research, various ML methods are used to determine which model is best for our traffic trace.

Chapter 3

Description of network sharing scenarios

3.1 Definition and Operational Framework of Network Sharing

Network sharing represents a strategic approach within telecommunication infrastructure management, wherein multiple base stations optimize resource utilization through cooperative operations. This entails the selective deactivation of specific base stations to leverage the capacity of neighboring units, thus achieving energy efficiency and reducing operational costs without significantly compromising quality of service.

3.2 Case Study: Operational Dynamics between Base Stations A and B

Consider base stations A and B as entities within this framework. The sharing dynamic between A and B is asymmetric; specifically, A and B's arrangement involves deactivating B to utilize A's operational capacity. This asymmetry arises due to the variable capacities inherent to each base station, which in turn affects the feasibility and efficiency of network sharing arrangements.

3.3 Network sharing distance potential analysis

In the context of optimizing the coverage of 4G cellular networks, the geometric coverage model of a base station is assumed to be circular, with a radius extending

up to 3 km. This model facilitates the examination of coverage overlap and network resilience, particularly in scenarios where redundancy is minimized without significantly compromising service availability. To achieve a coverage objective wherein 95% of users originally served by a pair of base stations remain within service range upon the deactivation of one station, a spatial analysis was conducted. The figure 3.1 indicate that pairs of base stations positioned no more than 472 meters apart satisfy this coverage criterion under ideal, unobstructed conditions. This spatial arrangement ensures that the operational base station can effectively cover 95% of the combined user base.

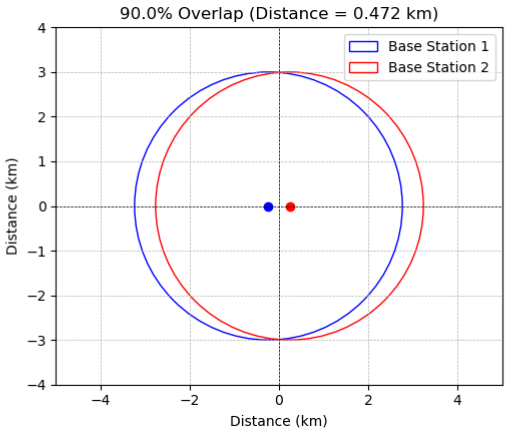

Figure 3.1: Analysising for the appropriate distance for network sharing

However, urban environments introduce complex variables that necessitate a departure from this idealized model. Factors such as building density, urban topology, and human activity concentrations significantly affect signal propagation, leading to the requirement for a denser network of base stations. Specifically, in city center areas, where obstacles are more prevalent and user density is higher, the analysis recommends a reduced separation distance between base station pairs. A threshold of 50 meters is proposed for these densely populated urban cores to maintain optimal service coverage. Conversely, for suburban areas, where user density decreases and line-of-sight conditions improve, a separation distance of 100 meters between base station pairs is deemed sufficient. This differentiated approach to base station deployment underscores the importance of tailoring network infrastructure to the specific characteristics of the serviced area. By adjusting the density of base stations in response to environmental and demographic factors, network operators can achieve a balanced trade-off between coverage efficiency and infrastructure cost.

3.4 Expected Outcomes

This thesis aims to contribute to the understanding and strategic planning of mobile network infrastructure, particularly focusing on the analysis of base station traffic traces and their geographical distribution. It aims to optimize network planning and facilitate infrastructure sharing for scenarios involving pairs of double and triple base stations. By implementing predefined strategies that set appropriate thresholds for switch operations, the study seeks to augment the practicality and efficiency of Network Sharing (NS) strategies. Furthermore, the research employs time series machine learning techniques to preemptively predict the utilization of network sharing, thereby enabling more informed and proactive network management decisions. This proactive approach is intended to ensure the consistent delivery of high-quality service to users.

Chapter 4

Workflow and methodology

4.1 Workflow

This study proposes a data-driven methodology to evaluate the efficiency of network sharing, focusing on energy conservation and financial savings across different urban and rural settings. The workflow is divided into several stages, each contributing to the overall goal of optimizing network resource utilization through strategic sharing arrangements between mobile network operators. The stages are as follows:

1. **Data Aggregation and Preprocessing:** The study first aggregates a month's worth of traffic load data, which includes 29 different categories that are gathered tile-by-tile. This detailed data collection facilitates a complete analysis of traffic distribution patterns, crucial for later processing phases.

2. **Selection of Mobile Network Operators and Geographic Areas:** The analysis targets base stations operated by two major telecommunications providers, Orange and Bouygues, across three demographically diverse regions: city centers, suburban areas, and rural locales. This selection ensures a comprehensive assessment of network sharing's potential across varied environments.

3. **Traffic Load Allocation via Euclidean Distance:** Using the Euclidean distance method, traffic loads for each tile are allocated to the nearest base station, yielding individual base station traffic traces. This approach reflects the unique operational footprint of each provider, resulting in operator-specific traffic profiles.

4. **Evaluation of Network Sharing Feasibility:** The feasibility of network sharing is evaluated by analyzing base station capacity and a preset threshold for sharing initiation. This determines the percentage of time slots that could implement network sharing while maintaining service quality.

5. **Energy and Economic Savings Calculation:** A predefined power consumption model estimates potential energy savings from network sharing. These savings are then translated into financial benefits, indicating the financial viability of network sharing.

6. **Expand the Strategy and Optimize Practicality:** The study explores an evolution from dual to triple base station (BS) network sharing configurations to enhance energy savings. Additionally, the frequency of switching operations, critical to the feasibility and practicality of network sharing strategies, is examined.

7. **Performance Metrics and Predictive Analysis:** A range of performance metrics is defined to assess network sharing efficiency. Machine learning techniques forecast future traffic patterns, improving operational efficiency by allowing proactive network sharing arrangements.

4.2 Methodology

4.2.1 Dataset

Overview

In this study, the research into base station (BS) uses the real normalized mobile traffic derived from the NetMob dataset [18], offered by a French mobile operator. This dataset represents a high-resolution cartography of multi-region, service-level mobile data traffic spanning 77 days. Data collection intervals occur every 15 minutes, featuring a spatial resolution of 100×100 square meters across 20 metropolitan areas in France. This huge amount of data captures traffic loads from more than 68 popular mobile services, covering a wide variety of digital activities such as social media engagement, video streaming, app downloads, web services, cloud computing, gaming, and music streaming. The dataset, accessible in text format via the official website, documents daily service-specific application data within a given region. Structurally, the dataset is organized such that the initial column records the tile number, indicative of the geographical location of a specific tile. Subsequent columns, totaling 96, enumerate the normalized traffic load for each 15-minute interval throughout the day. The example of the dataset is shown in figure 4.1. It is important to note that the row count varies by region, and due to the irregular boundaries of these areas, tile identifiers may not follow a consecutive sequence.

Complementing this dataset, the second source of data contains the geographical distribution of sites hosting mobile base stations. This includes detailed information on the mobile technology employed and the identity of the network operator (e.g.,

Bouygues, Free Mobile, Orange, SFR) associated with each base station. The example is shown in figure 4.2. Data for this component of the study are sourced from public datasets provided by the Agence Nationale des Frequences (ANFR) [19].

tile_id	time_point_1	time_point_2	time_point_3	time_point_4	time_point_5	time_point_6
66	7575	7210	5384	4802	7545	7111
353	7538	7278	5630	4848	7027	6843
354	6899	6464	4998	4561	6542	5969
640	5669	5677	4402	3834	5369	4806
641	6344	6193	4783	4440	6110	5264
642	12287	10032	8784	7470	11547	10698
643	11821	9408	8741	7448	11490	10647
926	2851	2902	2362	1972	2333	2328

Figure 4.1: Example of dataset of traffic load

Future studies will investigate network sharing policies based on actual data about base station locations and traffic loads in the area. In order to ensure a realistic representation of network demand, this calls for a methodical approach that aligns the allocation of traffic data from each specified block to its corresponding base station. Appropriate distribution of network load should be made based on the understanding of both the precise locations of the base stations and the regional traffic load data.

id;adm_lb_nom;sup_id;emr_lb_systeme;emr_dt;sta_nm_dpt;code_insee;generation;date_maj;sta_nm_anfr;na
1;SFR;513700;LTE 1800;2016-12-20;031;31118;4G;202 1.375555555556";43°47'36"N 1°22'32"E;En service
2;SFR;513700;GSM 900;2015-12-09;031;31118;2G;202 1.375555555556";43°47'36"N 1°22'32"E;En service
3;SFR;513700;LTE 700;2022-09-09;031;31118;4G;2023 1.375555555556";43°47'36"N 1°22'32"E;En service
4;SFR;513700;LTE 2100;2020-09-30;031;31118;4G;202 1.375555555556";43°47'36"N 1°22'32"E;En service
5;SFR;513700;LTE 2600;2018-07-13;031;31118;4G;202 1.375555555556";43°47'36"N 1°22'32"E;En service
6;SFR;513700;UMTS 900;2015-12-09;031;31118;3G;20 1.375555555556";43°47'36"N 1°22'32"E;En service
7;SFR;513700;5G NR 2100;2021-12-22;031;31118;5G;2 1.375555555556";43°47'36"N 1°22'32"E;Techniquen
8;SFR;513700;5G NR 3500;2022-12-21;031;31118;5G;2 1.375555555556";43°47'36"N 1°22'32"E;Techniquen
9;SFR;517600;LTE 1800;;045;45208;4G;2023-07-19;045 2.725";47°59'54"N 2°43'30"E;Projet approuvé
10;SFR;517600;UMTS 900;;045;45208;3G;2023-07-19;(2.725";47°59'54"N 2°43'30"E;Projet approuvé
11;SFR;517600;GSM 900;;045;45208;2G;2023-07-19;04 2.725";47°59'54"N 2°43'30"E;Projet approuvé
12;SFR;517600;LTE 800;;045;45208;4G;2023-07-19;045 2.725";47°59'54"N 2°43'30"E;Projet approuvé

Figure 4.2: Example of dataset of base stations

Data Process and preparation

Initially, Lyon was selected as the focal region for this study, chosen for its moderate size, which accurately reflects the urban dynamics of a typical city. The preliminary step involved translating the dataset's tile data into column and row indices to

determine the precise locations and corresponding traffic loads within the region. Formal dimension data revealed that Lyon covers a total of 426 rows and 287 columns. Through this conversion process, Figure 4.3 was generated, illustrating the distribution of YouTube traffic volume across Lyon at a specific time point. This visualization employs a color gradient, with warmer areas such as red indicating higher normalized traffic volumes, while cooler shades denote regions with lower traffic activity. Figure 4.3 shows that traffic volumes vary across different areas, with notably higher traffic loads in central locations. Regarding the base station data, an initial filtering step was to isolate in-service 4G base stations operated by the four main French telecommunications providers: Orange, SFR, Bouygues Telecom, and Free Mobile. This filtration was essential to streamline the focus of the study. The resulting data encompassed base stations within the geographical bounds of Lyon, defined by a longitude (x) range of approximately 4.6924 to 5.0603 and a latitude (y) range of approximately 45.5571 to 45.9392. A computational formula was subsequently developed to align the real-world positions of these base stations with the tile index system used in the dataset. The latitude span is 0.3821 which represents 287 columns and the longitude span is 0.3679 which represents 426 rows. The conversion of tile numbers to their corresponding row and column indices is achieved through the following operations:

- The `row_index` is obtained by taking the integer division of `tile_id` by the number of columns (`n_cols`):

$$\text{row\_index} = \text{int}\left(\frac{\text{tile\_id}}{\text{n\_cols}}\right)$$

- The `col_index` is determined by computing the remainder of `tile_id` divided by `n_cols`, effectively capturing the integer remainder of the division:

$$\text{col\_index} = \text{int}\left(\text{tile\_id} \mod \text{n\_cols}\right)$$

For the base stations, conversion from geographical coordinates to row and column indices is achieved as follows:

- The `row_index` for base stations is calculated based on the latitude:

$$\text{row\_index} = \left(\frac{\text{base station latitude} - 45.5571}{0.3821/287}\right)$$

- The `col_index` for base stations is calculated based on the longitude:

$$\text{col\_index} = \left(\frac{\text{base station longitude} - 4.6924}{0.3679/426}\right)$$

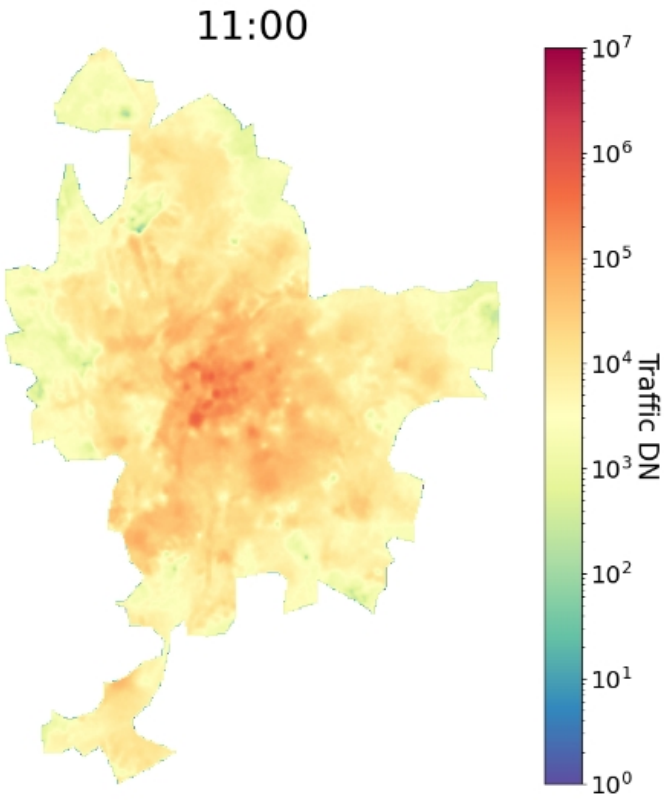

Figure 4.3: YouTube traffic volume across Lyon at a specific time point

After this synchronization of location definitions proceed a clear understanding of the base station distribution relative to the city's tiles, laying the groundwork for more detailed subsequent analyses.

Modeling the raw traffic traces

To realize the concept of network sharing within base station infrastructure, the foundational step involves acquiring the traffic trace data for each base station. The traffic volume data we possess is characterized by a spatial resolution of 100×100 square meters. However, in practice, not every spatial tile may house a base station, and in densely populated areas such as city centers, a single block might contain multiple base stations. This uneven distribution of base stations requires the adoption of more sophisticated methods to accurately assign traffic data to specific base stations.

Under ideal conditions with no physical obstacles, the distance between a block and a base station—which is primarily defined by Euclidean distance—becomes an important consideration during the allocation procedure.

To address this, we employ an Euclidean distance-based method to efficiently allocate traffic volumes to the nearest base station. Specifically, we aim to map the traffic volume traces, spread across an $n \times n$ grid of tiles, to the respective base stations, denoted by Op_1, within the area of Lyon.

Let $L = \{L_1, L_2, \ldots, L_{n \times n}\}$ represent the set of traffic time series for the specified area, and $B = \{B_1, B_2, \ldots, B_m\}$ the set of base stations operated by Op_1. The traffic volume corresponding to each time series L_i from tile i is allocated to the nearest base station B_j, minimizing the Euclidean distance d_{ij}. Figure 4.4 illustrates the mapping of traffic volumes from each tile to the base stations within a given urban area at a specific time interval, with the intensity of color indicating the level of traffic volume in the corresponding tile for a sample time slot.

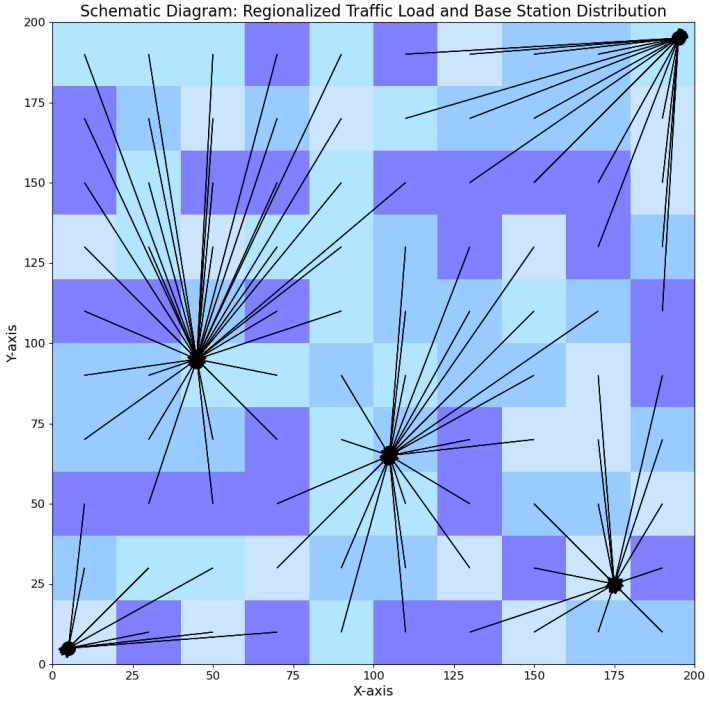

Figure 4.4: "Euclidean method" for distributing block traffic load to BSs

Finally, the actual traffic trace managed by each base station B_i is evaluated through the aggregation of traffic data series allocated to it via the aforementioned strategy.

4.2.2 Network Sharing strategies

Threshold Determination for Network Sharing

A critical aspect of implementing network sharing is the establishment of a traffic load threshold which is C_{th}, set at 80% for this analysis most of time. 80% is a conservative number that will not overload the base station. C_{th} stipulates that network sharing becomes viable when the sum of offloaded traffic from a deactivated base station (e.g., A) and the traffic from the active origin base station does not exceed 80% of the capacity of the receiving station (e.g., B). If the cumulative traffic load passes C_{th}, both base stations must remain operational to fulfill the demand requirements adequately. Then the Analysis of Threshold Variability is researched to investigates the impact of varying C_{th}, with 5% intervals ranging from 50% to 80%, on the proportion of achievable network sharing and associated reductions in power consumption. This analysis elucidates the relationship between C_{th} settings and network sharing efficiency.

BS Capacity Definition and Optimization

A notable challenge in this domain is the accurate definition of base station capacity, particularly given that available traffic volume data are normalized rather than real-world quantities. It was not practical to define capacity based on 90% of the maximum traffic volume that was observed at first. This was because unusual traffic peaks caused capacity estimates to be significantly overvalued sometimes up to six times the average load. Regarding the design capacity of base stations to accommodate peak traffic demands, a factor of six times the average load is often cited as a benchmark for robust network design. However, for the purposes of this research, a more conservative approach is adopted to estimate the requisite capacity. Therefore, a refined methodology involves identifying the 10th highest traffic volume as a more representative peak and calculating 90% of this value to determine base station capacity. This adjustment mitigates the influence of extreme outliers, facilitating a more pragmatic and reliable framework for network sharing parameterization. In our study of base station pairs engaged in network sharing, we use a strategy in which the station with lower capacity is consistently deactivated, supposing the station with higher capacity is active to ensure longer network sharing and larger energy savings. This NS status framework is quantified using a binary status indicator: '1' signifies the feasibility and activation of network sharing, whereas '0' indicates periods when network sharing is not operational. This binary representation facilitates clearer research analysis and simplifies calculations, allowing for an intuitive examination of network sharing's efficiency and sustainability impact.

Area definition

In the context of city planning and telecommunications infrastructure, the spatial distribution of base stations and population density across various city sectors are not the same. Consequently, the demand for network bandwidth and traffic volume exhibits considerable variability, underscoring the potential region-specific research to inform network sharing strategies. Commercial districts, for example, are likely to experience peak traffic volumes during working hours and lunch breaks, attributable to the high concentration of economic activities and employee presence. Conversely, residential zones typically see a surge in network usage during evening hours, reflecting the population's engagement in leisure and communication activities after work hours. Mixed-use areas, on the other hand, may demonstrate a blend of these patterns or maintain a relatively consistent level of network usage throughout the day. The variability extends to different days of the week as well; weekdays are characterized by traffic loads concentrated around workplaces during business hours, while weekends witness increased network activity in residential and commercial zones, driven by leisure and entertainment purposes. Therefore, I propose a categorization of the Lyon metropolitan area into three distinct zones based on base station density. This classification serves as a foundational framework for subsequent case studies,

- **City Center:** Characterized by a base station density exceeding 3 units per square kilometer.
- **Suburban:** Defined by a base station density ranging from 1.5 to 3 units per square kilometer.
- **Rural:** Identified by a base station density less than 1.5 units per square kilometer.

Based on these three regions' constraints, the subsequent case study looks into whether there are any variations in the network sharing outcomes for them.

Methodological Approaches for Evaluating Power consumption of base station

This section focuses on the methodological framework adopted for assessing the power consumption dynamics of base stations, specifically focusing on Remote Radio Heads (RRHs). The analysis is grounded in the principles outlined in the paper [20] which posits a near-linear relationship between the Radio Frequency (RF) output power (P_{out}) and the base station's power consumption (P_{in}).

In the context of RRHs, the power consumption model is articulated as follows:

$$P_{\text{in}} = N_{\text{TRX}} \times (P_0 + \delta p \times P_{\text{out}})$$

where:

- P_{out} denotes the relative RF output power, constrained by $0 < P_{\text{out}} \leq P_{\max}$,
- P_{out} is directly proportional to the traffic load-to-capacity ratio of the base station,
- N_{TRX} represents the number of transceiver modules,
- P_0 is the power consumption at zero RF output power (a fixed linear model parameter),
- δp signifies the slope of the load-dependent power consumption.

Given the specifications for RRH, with $N_{\text{TRX}} = 6$, $P_{\max} = 20W$, P_0, and $\delta p = 2.8$, the power output (P_{out}) is mathematically represented by the ratio of the traffic load to the base station's capacity, scaled by $P_{\max}$:

$$P_{\text{out}} = \left(\frac{\text{traffic load}}{\text{capacity of base station}}\right) \times P_{\max}$$

Comprehensive Power Consumption Model

Incorporating these parameters, the comprehensive model for calculating the base station's power consumption emerges as:

$$P_{\text{in}} = 504 + 16.8 \times \left(\frac{\text{traffic load}}{\text{capacity of base station}}\right) \times 20$$

This formula offers a precise quantification of energy usage based on fluctuating traffic loads and the inherent capacity of the RRH base station. Subsequently, the model serves as a primary tool for evaluating the energy savings attributable to network sharing. By comparing the power consumption under scenarios involving network sharing strategy against standard operational scenarios, a respectable result could be obtained to highlight the potential for significant energy efficiency benefits within telecommunications infrastructure sharing.

4.2.3 Switch operation

Switch operation definition

The concept of switch operation can be alternatively described as changes in NS status, which is predetermined as a binary value—1 or 0. Here, a status of 1 indicates that network sharing is active during a given time period, implying that one base station remains operational while the other is deactivated. Conversely, a status of 0 denotes that both base stations are active. A transition from 1 to 0 or vice versa is counted as a single switch operation.

Switch efficiency definition

Switch efficiency is identified as a critical metric in NS, illustrating the importance of evaluating switch operations not only by quantity but also by their impact on NS utilization. This metric indicates how many minutes that a single state change may provide for NS. Ideally, each switch operation should maximize the amount of NS time it brings.

The analytical formula is shown as follow:

$$\text{Switching efficiency} = \frac{\text{period\_share} \times 15\text{min}}{\text{switch state change}}$$

4.2.4 Traffic prediction

LSTMs introduction

Long Short-Term Memory (LSTMs) networks are a special kind of Recurrent Neural Network (RNN) capable of learning long-term dependencies in data sequences. Unlike traditional RNNs, which struggle to capture long-term temporal relationships due to the vanishing gradient problem, LSTMs are designed with a unique mechanism. This mechanism allows them to effectively remember information over extended periods and forget irrelevant data, making them particularly suited for time series prediction tasks where understanding past context is crucial. In the context of traffic load prediction, LSTMs offer several benefits. Firstly, their ability to process sequential data makes them ideal for analyzing time series where the sequence of data points is significant. Traffic patterns are naturally sequential and have temporal dependencies, which LSTMs can detect and model. Second, LSTMs can deal with variability and seasonality in traffic data, learning from daily, weekly, and even yearly patterns to make accurate predictions. Finally, their ability to handle different lengths of input sequences enables the modeling of traffic loads with varying historical data lengths, as well as accommodating sudden changes in traffic patterns caused by special events.

XGBoost introduction

XGBoost, short for Extreme Gradient Boosting, is an advanced and optimized gradient boosting framework that is skilled at regression and classification. It stands out for its efficiency, flexibility, and cross-platform compatibility. XGBoost improves prediction accuracy by sequentially building decision trees, with each tree attempting to correct the mistakes of its predecessors, thereby effectively reducing loss and improving results. When applied to traffic load prediction, XGBoost proves advantageous due to its: effortlessly handling missing values and discrepancies, ensuring accurate predictions even with incomplete datasets.

Its ability to perform feature importance analysis aids in understanding and ranking the factors that influence traffic load, allowing for the development of targeted management strategies. XGBoost's speed-focused design allow it to efficiently process large datasets, making it ideal for real-time traffic load prediction. Furthermore, XGBoost uses regularization to prevent overfitting, allowing the model to generalize to previously unseen data while ensuring its robustness and forecasting accuracy. XGBoost's precision and efficiency in learning from complex traffic data sets make it a powerful tool for forecasting, facilitating effective traffic management and resource optimization.

ARIMA introduction

The ARIMA (Autoregressive Integrated Moving Average) model is a classic statistical technique for forecasting time series data, combining autoregression (AR), differencing (I), and moving average (MA). It's highly regarded in traffic load prediction for its ability to capture the dependencies between observations, handle seasonal variations, and offer flexibility by adjusting its parameters (p, d, q). Despite its simplicity, ARIMA excels in producing accurate forecasts, especially for well-understood time series, making it ideal for analyzing traffic patterns. Its emphasis on stationarity through data differencing not only ensures precise short-term forecasts but also facilitates a deeper analysis of traffic data. This makes ARIMA a valuable tool for improving traffic management and optimizing network resources.

Chapter 5

Traffic load and base station characterization

5.1 Traffic load analysis

The NetMob dataset comprises data from 68 widely utilized mobile services, representing a broad spectrum of applications fundamental to modern life. These include offerings from major platforms such as the Apple App Store, Facebook, Google's suite of applications, Netflix, YouTube, Spotify, and the PlayStation network. For this analysis, a focused subset of 29 applications was chosen as table 5.1 shows. These applications were subsequently classified into distinct traffic load types, namely video streaming, social media, application store downloads, web-related services, cloud services, gaming, and music streaming. This classification aimed to enable a nuanced examination of traffic distribution across different types of digital services in daily use.

Amazon Web Services	Apple App Store	Apple iCloud	Apple Music
Apple Video	DailyMotion	Dropbox	EA Games
Facebook	Facebook Live	Fortnite	Google Drive
Google Meet	Microsoft Store	Netflix	Orange TV
PlayStation	Pokemon GO	Skype	Spotify
Telegram	Twitch	Twitter	Web Downloads
Web Games	Web Streaming	WhatsApp	Wikipedia
YouTube			

Table 5.1: List of Applications

The findings from figure 5.1 reveal that video streaming services account for

the majority of data traffic, representing over half of total consumption. This is followed by significant contributions from social media platforms and application stores, highlighting the extensive role of social media Subsequent research sought to analyze peak and off-peak usage times for the two dominant traffic categories, investigating whether usage patterns for these application types exhibit different cycles. The figure 5.2 indicated that social media usage peaks around 13:00-13:15, aligning with typical lunch break times, suggesting a tendency for individuals to engage with the internet post-lunch. Conversely, video streaming activities figure 5.3 peak between 22:00-22:45, reflecting common pre-sleep routines where individuals consume TV programs or movies as a form of relaxation.

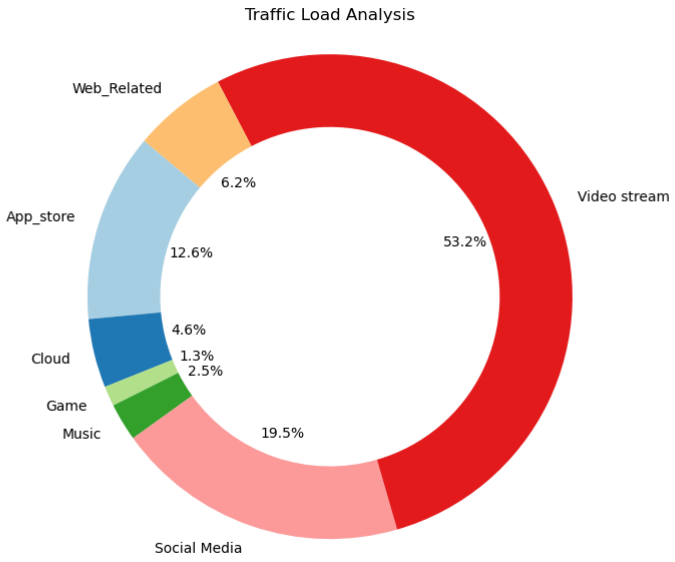

Figure 5.1: Percentage of different types of traffic load

The analysis also identified a uniform off-peak period for both application types between 04:00-06:00, a period when most people are asleep and thus digital activity is minimized. This period is ideal for putting network sharing strategies into practice, taking advantage of the decreased demand to maximize the effectiveness and allocation of network resources.

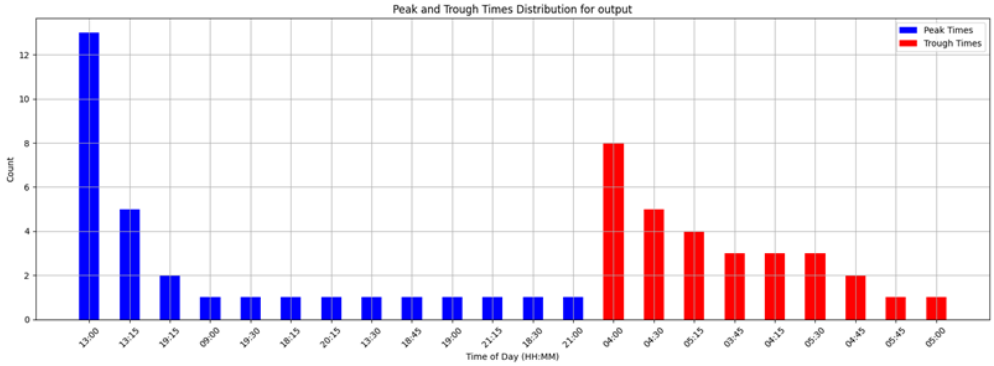

Figure 5.2: Social media traffic pattern analysis

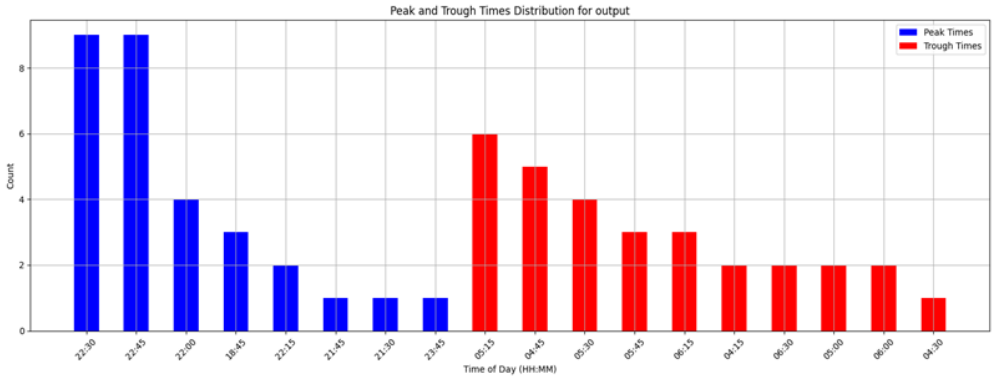

Figure 5.3: Video streaming traffic pattern analysis

5.2 BS traffic trace analysis

After analyzing the traffic profiles for each base station (BS) operated by Operator 1 (Op_1), a specific interest was developed concerning the traffic load patterns of Orange base stations. This investigation employed the K-means clustering method to categorize the base stations based on traffic volume, without prior knowledge of the cluster count. The initial step involved determining the optimal number of clusters (K) using the Elbow method, which identifies the "elbow" point where the rate of decrease in the Within-Cluster Sum of Squares (WCSS) begins to diminish. The analysis figure 5.4 suggested that $K = 4$ is an appropriate choice.

Subsequently, Orange base stations were segmented into four distinct clusters, with the traffic volume serving as the classification criterion. Notably, the average traffic volume for base stations in each successive cluster $(i + 1)$ was approximately ten times greater than that of the preceding cluster (i). Table 5.2 presents the observed count of base stations in each cluster alongside the normalized average

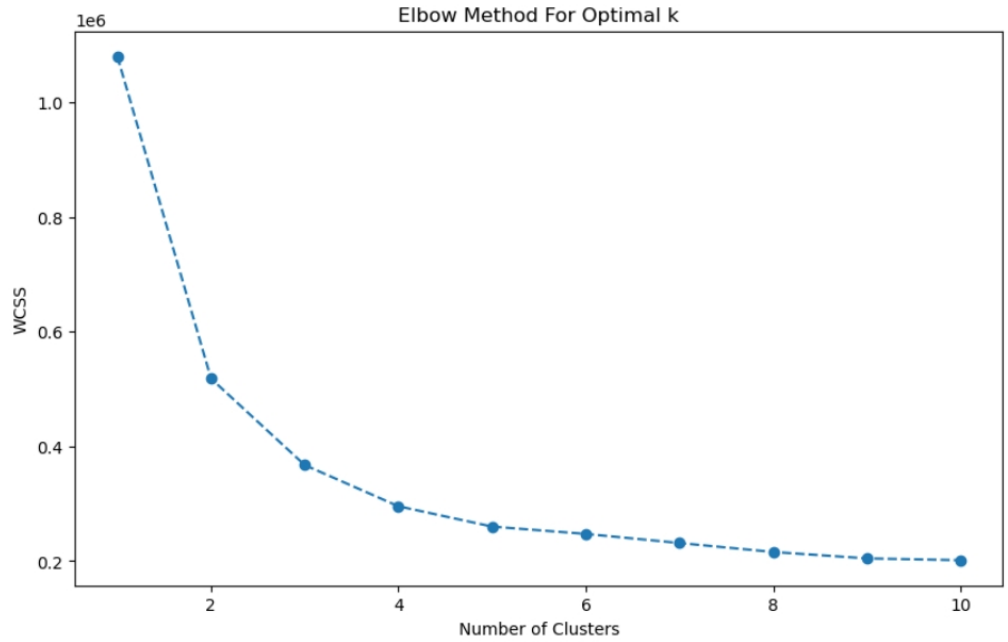

Figure 5.4: Elbow method for determining K

traffic volume per base station. Furthermore, Figure 5.5 illustrates the spatial distribution of base stations across Lyon, differentiated by cluster.

Cluster ID	**0**	**1**	**2**	**3**
Number of BSs	265	118	34	8
Normalized traffic volume per BS	0.31	3.33	30.91	288.63

Table 5.2: BS clusters number and normalized volumn

The analysis revealed that clusters 2 and 3 comprised a smaller number of base stations, each handling high to extremely high traffic volumes and dispersed sparsely across the region. This pattern indicates a relatively low potential for network sharing due to the higher demand at these nodes. Conversely, a significant number of base stations were classified under Cluster 1, characterized by lower traffic volumes. This cluster, coupled with a denser distribution of base stations that likely results in overlapping coverage areas, suggests a higher feasibility for implementing network sharing strategies. The greatest potential for network sharing was identified within Cluster 0, which consists of many base stations experiencing extremely low traffic loads. These stations are densely situated, predominantly in urban centers, to supplement capacity during periods of peak demand, thereby indicating optimal conditions for network sharing.

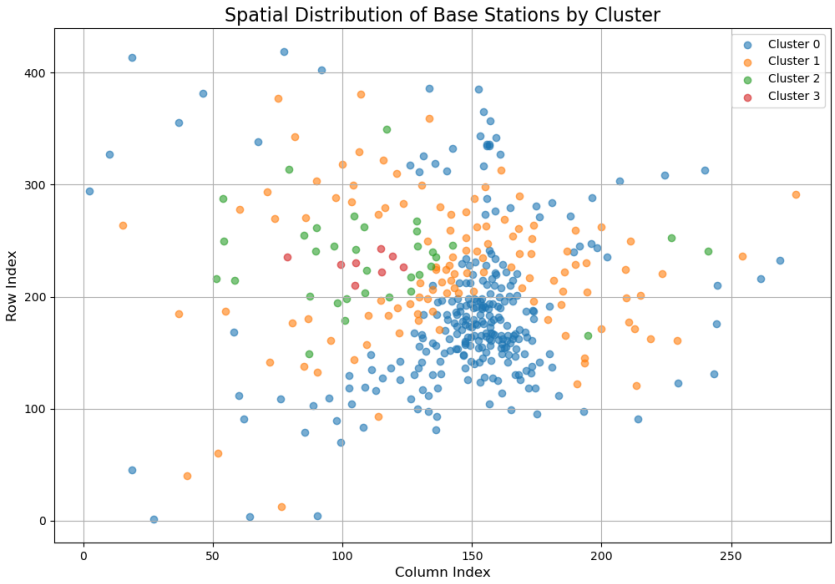

Figure 5.5: The spatial distribution of base stations across Lyon in different clusters

5.3 Base station distribution analysis

This part I provide a preliminary evaluation of the network sharing potential analyzing how the BSs from different mobile operators are distributed in the city of Lyon

5.3.1 Nearest Neighbor Analysis

This method measures the distances between each point and its nearest neighboring point. A smaller average distance indicates a more clustered pattern. Table 5.3 reports the average intra-operator BS distance, corresponding to the average distance between each BS and the nearest BS belonging to the same operator. A lower average distance reflects a more clustered pattern of BS distribution, possibly indicating a focus of the mobile operator on a denser BS deployment in the Lyon region, like in the case of SFR.

	Orange	**Bouygues Telecom**	**SFR**	**Free Mobile**
Distance [m]	791	733	697	832

Table 5.3: Intra-operator BS distance

5.3.2 Huff Model

The Huff Model will help us understand the probability of the base stations of one operator to be located near the base stations of another operator. Table 5.4 shows the average inter-operator Huff Model probability for spatial interactions between base stations of different operators, reflecting the likelihood that the BSs owned by a given operator are located near the BSs belonging to another operator. The Huff Model is typically adopted in spatial analysis to evaluate the probability for a user to visit a site, based on the distance of the site, its attractiveness, and the relative attractiveness of alternatives.

	Huff Model Probability
Orange-Bouygues Telecom	0.473
Orange-SFR	0.463
Orange-Free Mobile	0.558
Bouygues Telecom-SFR	0.492
Bouygues Telecom-Free Mobile	0.531
Free Mobile-SFR	0.505

Table 5.4: Huff Model probability for all pairs of MOs.

Based on the Huff Model, the probability that a BS at location x_i will interact with any BS in set Y due to its closeness The normalized Huff Model Probability, $P(x_i \to Y)$, the probability that a BS at location x_i will interact with any BS in set Y due to its closeness. For each base station x_i in set X is calculated as follows:

$$P(x_i \to Y) = \frac{M_1}{\sum_{j=1}^{M_2} \left(\frac{M_2}{(d_{ij}+1\times10^{-10})^{\lambda}} \right) + M_1} \tag{5.1}$$

where:

- M_1 and M_2 are the number of base stations for each operator.
- d_{ij} represents the Euclidean distance between base station i from set X and base station j from set Y.
- λ is a given sensitivity to distance parameter.

To calculate the mean Huff Model Probability, P_H, for all base stations in set X interacting with any base station in set Y, the formula is:

$$P_H = \frac{1}{R} \sum_{i=1}^{R} P(x_i \to Y) \tag{5.2}$$

where R is the number of base stations in set X. The pair Orange-Free Mobile features the highest Huff Model probability, suggesting a stronger spatial closeness between BSs from these two operators, possibly highlighting a higher potential for network sharing that may derive from the cooperation between these two operators in this region.

5.3.3 Ripley's K-function and G-function

From the paper "Characterizing Spatial Patterns of Base Stations in Cellular Networks point analysis", Ripley's K-function and G-function are utilized to determine whether the base station distribution is uniform or exhibits a clustered pattern. These two coefficients enable the analysis of base station distribution across various areas. Different operators exhibit distinct coefficients, introducing variability into these measures, and necessitate definitions for urban and rural areas.

Ripley's K-function

Ripley's K-function is a statistical tool used in spatial analysis to quantify the degree of spatial clustering or dispersion of points within a specified area. It serves as a measure of spatial homogeneity and provides insights into the patterns of distribution—whether points are randomly dispersed, uniformly spaced, or clustered—over a range of scales.

The mathematical expression for Ripley's K-function, denoted as $K(t)$, for a set of points within a bounded region A is given by:

$$K(t) = \lambda \sum_{i \neq j} \frac{I(d_{ij} < t)}{|A_{ij}|}$$

where:

- t represents the scale of analysis or the distance of interest.
- λ is the intensity of the point process, defined as the average number of points per unit area.
- d_{ij} is the distance between points i and j.
- I is the indicator function that equals 1 if the condition within its argument is true (i.e., $d_{ij} < t$) and 0 otherwise.

- $|A_{ij}|$ accounts for edge effects by adjusting the contribution of each pair of points based on their proximity to the boundary of the region A.

Ripley's G-function

Ripley's G-function is a significant statistical tool that analyzes the spatial patterns by focusing on the nearest-neighbor distance distribution within a point pattern.

The G-function, $G(d)$, can be defined as the cumulative distribution function (CDF) of the nearest-neighbor distances in a point pattern. Mathematically, it is expressed as:

$$G(d) = P(D \leq d)$$

where D represents the nearest-neighbor distance for a randomly selected point, and d is a specific distance of interest. The value of $G(d)$ indicates the proportion of point pairs whose nearest-neighbor distance is less than or equal to d.

The interpretation of the G-function revolves around understanding the spatial closeness of points within the pattern. A higher $G(d)$ value at a particular distance d suggests a greater level of clustering or aggregation at that scale, as more points have their nearest neighbors within this distance. In contrast, lower $G(d)$ values indicate a more dispersed or regular distribution, with points being farther apart from their nearest neighbors.

A larger K-function value suggests a more clustered distribution of base stations in city centers, as shown in figure 5.6. Because of the increased demand for mobile network services brought on by larger populations and more concentrated urban activities, it appears that base stations are distributed more widely. Furthermore, the G-function for cities shows a sharp increase in the nearest neighbor clustering. Simultaneously, the G-function for the rural area gradually increases, suggesting a more uniform distribution of nearest neighbors.

Implications for Network Sharing: The spatial distribution shown by Ripley's K-function may have an impact on the viability and possible advantages of sharing infrastructure in network sharing strategies. Closely spaced base stations from various operators could make sharing arrangements easier and more advantageous in city centers, which could result in significant cost savings and efficiency benefits. The viability of sharing in rural areas may be influenced more by geographical positions because base stations are more widely distributed there.

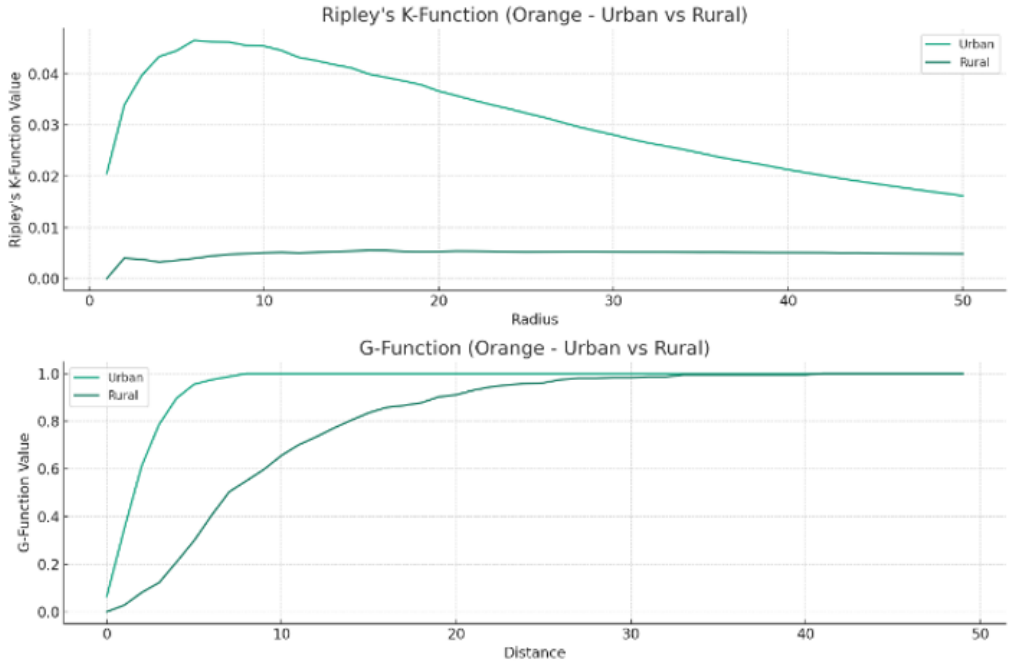

Figure 5.6: The Ripley's K-function and Ripley's G-function of Orange BSs in urban area and rural area

Chapter 6

Double base stations network sharing

6.1 Double Base Station Network Sharing Performance Evaluation in Different Areas

This part of the study concentrates on where base stations (BS) are close to each other, operated by two mobile operators (MOs), namely Orange (Op1) and Bouygues Telecom (Op2). These operators are scrutinized across three distinct geographic contexts: urban, suburban, and rural settings which were mentioned before, each offering unique operational challenges and opportunities for network sharing (NS). Utilizing time-variant BS traffic trace data which is integrated before, the research first quantifies the fraction of time slots amenable to NS implementation, symbolized as f_T. The critical variable under investigation is the threshold C_{th}, which delineates the maximum allowed saturation of BS capacity for offloading traffic to initiate NS. The dependency of f_T on varying C_{th} values is meticulously analyzed through traffic patterns observed across different days of the week, highlighting the temporal dynamics of NS applicability. The findings in Figures 6.1 and 6.2, with each curve representing a different day of the week (with Monday corresponding to Day 1), illustrate a noteworthy positive correlation between the f_T and the threshold C_{th}, with higher thresholds expanding the scope of NS applicability, with the f_T surpassing 90%. A notable disparity in f_T values is observed between weekdays and weekends, especially in urban and suburban locales, likely due to high traffic demands on these areas on weekdays. This variance underscores the potential of NS to adapt to changing network demands, as opposed to weekends where NS strategies are feasibly employed for extended durations, even under minimal C_{th} configurations.

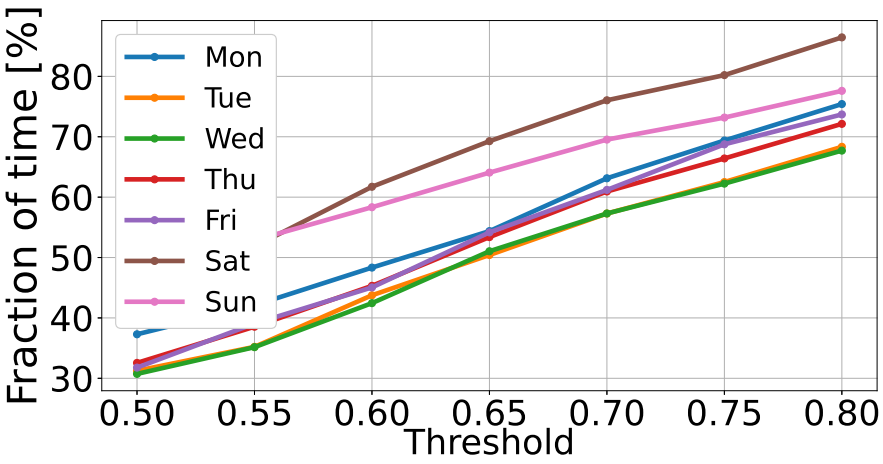

Figure 6.1: Fraction of NS time slots in urban area

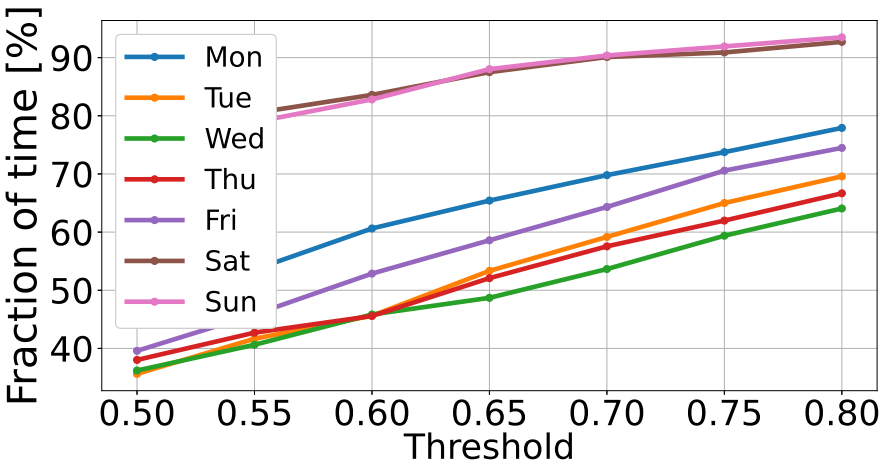

Figure 6.2: Fraction of NS time slots in suburban area

In contrast, rural areas, as shown in Figure 6.3, display minimal daily variability, with no obvious distinction between weekdays and weekends. Remarkably, even with C_{th} thresholds as low as 50%, NS application extends to at least one-third of the day across all regions, peaking at nearly 80% in suburban areas. These findings underscore the significant leeway mobile operators possess in mitigating network capacity overload risks. Conservative C_{th} settings still permit the deactivation of superfluous radio resources for considerable periods, presumably leading to substantial reductions in energy consumption and operational expenditures. Moreover, the analysis reveals a substantial overlap in the time slots during which either of

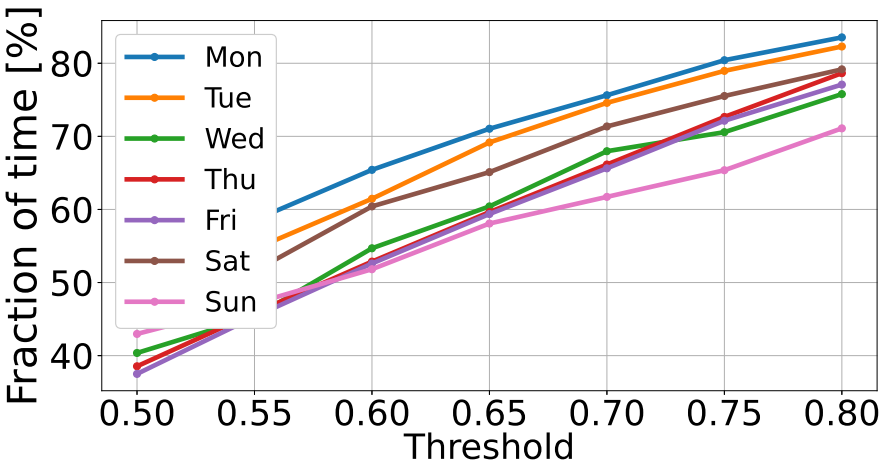

Figure 6.3: Fraction of NS time slots in rural area

Figure 6.4: Fraction of NS time slots considering three different area types.

the co-located BSs within each pair might be deactivated, suggesting periods or usage patterns where NS is broadly feasible.

Furthermore, the study investigates the implications of C_{th} settings on the frequency of BS activation and deactivation operations. Figure 6.8 depicts the occurrence of BS activation/deactivation operations for increasing values of C_{th} during the different days of the week in the three area types. Frequent BS operations will reduce the life of the switch and make network sharing inconvenient for operators. Contrary to f_T's behavior, this metric exhibits greater variability and lacks a consistent pattern across different days, A higher operation suggests that traffic patterns may include either multiple short low-traffic intervals or fewer longer off-peak periods. Notably, significant f_T achievements come at the cost of frequent BS switching operations, especially in urban areas, raising concerns over operational acceptability due to high switching frequencies. These critical aspects should be taken into account in defining the proper threshold configuration, since it may be convenient to limit the maximum allowed frequency of switching operations or to introduce some hysteresis by setting the threshold on the BS operations, to avoid too frequent switching operations, hence preserving the BS from faster degradation. This is done in the following chapter research.

Furthermore, another research is to focus on energy consumption. Figure 6.12 shows that the strategic application of NS could facilitate substantial energy savings, potentially exceeding 40% compared to scenarios of no NS implementation. The results illustrate the empirical findings related to energy conservation achieved

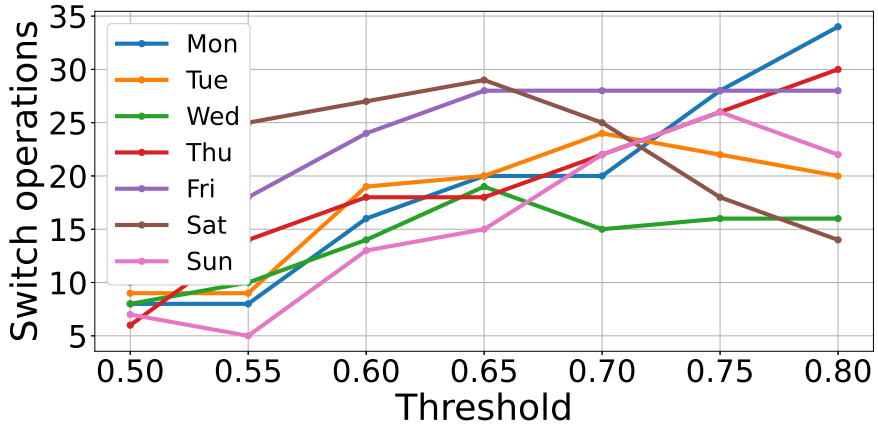

Figure 6.5: Switch operations in urban area

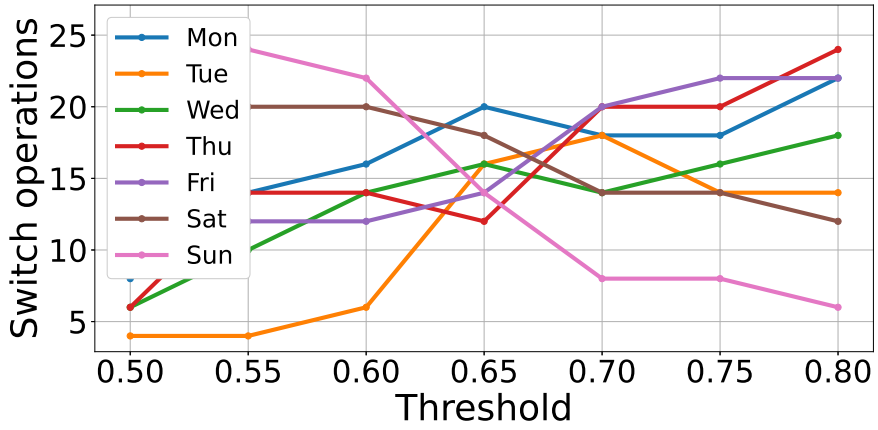

Figure 6.6: Switch operations in suburban area

through NS across various geographic settings, employing a representative sample of base station (BS) pairs. The analysis delineates the correlation between energy savings and the C_{th} settings, applied uniformly across each day of the week. It is observed that elevating the C_{th} threshold contributes to increased energy savings, albeit with notable fluctuations dependent on daily traffic loads and user interaction patterns. The figure for the fraction of NS time slots and the figure of energy saving shows the same pattern which means that there may be a linear correlation between the energy saved and f_T.

In rural areas, the disparity in energy conservation between weekdays and weekends is not obvious, suggesting a uniform potential for savings each day. Remarkably, even under a low C_{th} configurations, a minimum energy saving of 10%—and in some cases, up to 35%—can be realized across all examined areas. This

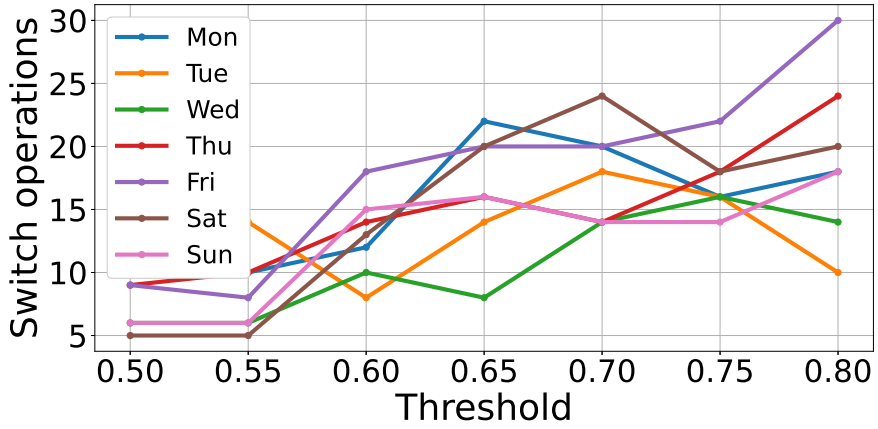

Figure 6.7: Switch operations in rural area

Figure 6.8: Switch operations considering three different area types.

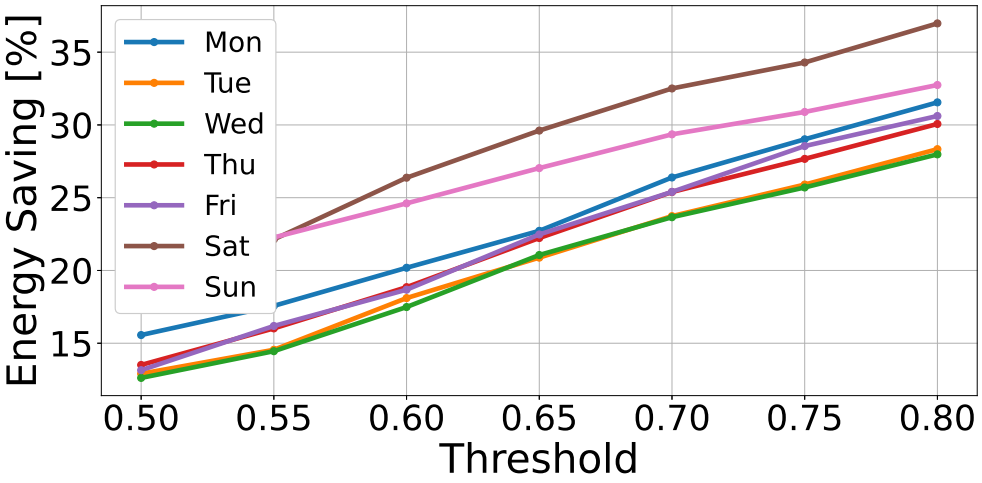

Figure 6.9: Energy saving in urban area

finding emphasizes the capability of NS to balance the objectives of a sustainable green network with the maintenance of the quality of service (QoS).

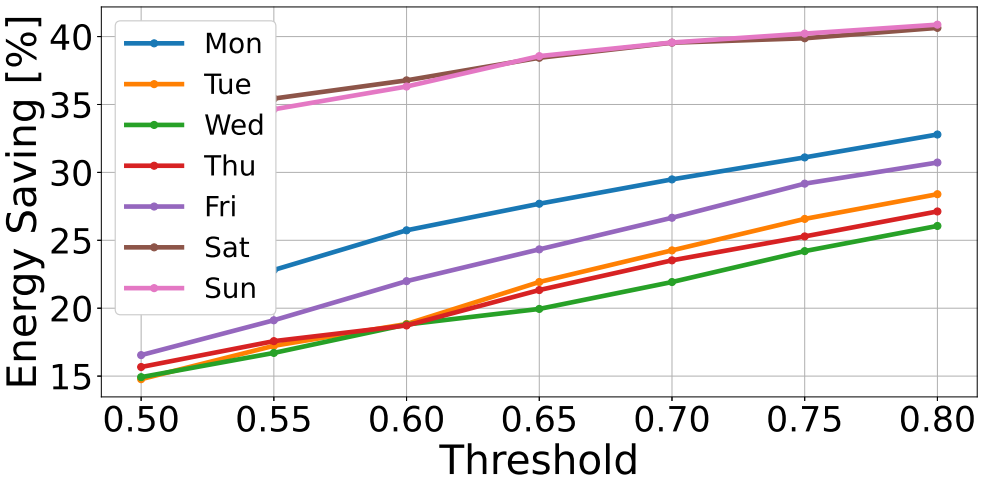

Figure 6.10: Energy saving in suburban area

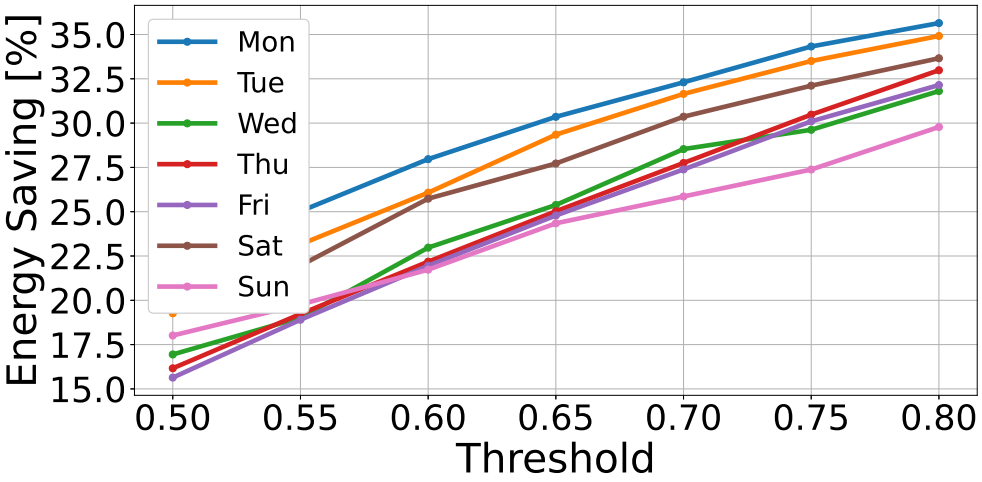

Figure 6.11: Energy saving in rural area

Figure 6.12: Energy saving considering three different area types.

Chapter 7

Network sharing performance metrics

In the domain of network sharing, it is essential to conduct the evaluation of its performance outcomes, focusing on aspects such as energy savings, the feasibility of implementing sharing configurations across base station pairs, and the optimal frequency of base station switch operations. This kind of analysis is essential for defining the real-world advantages and operational sustainability of network sharing in the context of managing telecommunications infrastructure. Furthermore, it becomes crucial to carefully examine how frequently infrastructure switch operations occur. The idea behind these operations in practical application is to minimize their frequency as much as feasible. This method is motivated by the knowledge that excessive switching might cause base station components to start degrading more quickly, which would be harmful to the network infrastructure's resilience and operational lifespan. Here are the detailed analysis of different metric parameters:

7.1 Base station remaining percentage

This metric examines the utilization patterns of active base stations within network sharing configurations, aiming to delineate the daily usage trends of base station resources. The formula for calculating the remaining percentage of base station capacity at each time point is defined as:

$$\text{Base Station Remaining Percentage} = \frac{\text{capacity-tf}_{\text{curr}}}{\text{capacity}}$$

Where "Capacity" refers to the total capacity of the active base station, and "Current Traffic Flow" utilizes data from the operational base station within a network sharing pair.

Analysis of the figure 7.1 and figure 7.2 yields distinct patterns of base station capacity utilization across different regions. In urban centers, the remaining capacity of base stations during weekdays is notably low, with a marked increase over weekends. This pattern suggests that base stations in these areas predominantly serve workplaces, experiencing peak traffic loads after lunch, a trend that aligns with prior analyses focusing on social media traffic loads. In such contexts, social media traffic constitutes a significant portion of the total network load.

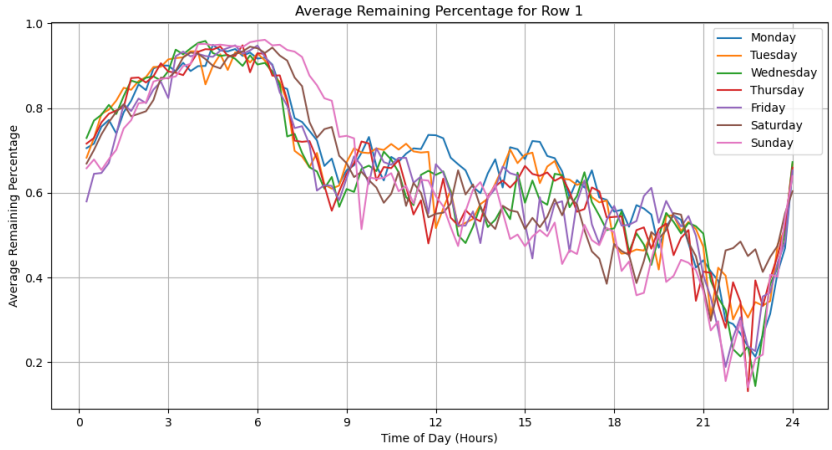

Figure 7.1: Base station remain percentage for rural area

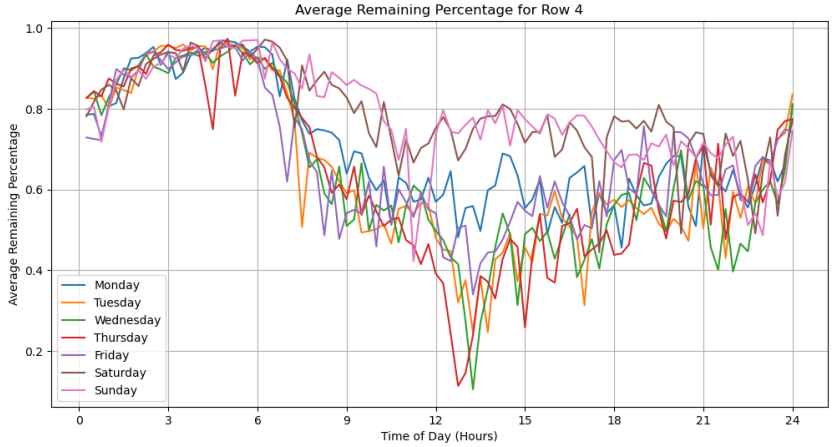

Figure 7.2: Base station remain percentage for city center area

Conversely, rural regions exhibit a different pattern, characterized by minimal

fluctuations in remaining capacity during morning and afternoon hours, and a sharp decrease in the evening as people engage in home-based activities. This observation aligns with previous findings, suggesting that rural areas, likely serving residential communities, experience traffic patterns influenced by evening activities, such as streaming videos or communicating with social media for entertainment and information. This indicates that base stations in residential areas face increased load pressures during evening hours. All these observations underscore a close relationship between base station usage patterns and the daily life activities of individuals in different regions. In order to account for these variations, it is crucial to dynamically adjust network sharing strategic arrangements.

7.2 Traffic load transfer

The concept of traffic load transfer is quantified as the proportion of network load that is offloaded from a deactivated base station to an active counterpart within a network sharing arrangement. This metric serves as an indicator of the offloading capability inherent to a pair of base stations. The study embarks on a comparative analysis across various geographic regions and among different network operators to ascertain potential correlations between the traffic load transfer rate and the specific characteristics of regions or operators.

The analytical model for calculating the traffic load transfer is expressed by the formula:

$$\text{Traffic load transfer} = \frac{\sum_{i=1}^{n} \text{tf}_i \cdot \text{sharing\_status}_i}{\sum_{i=1}^{n} \text{tf}_i}$$

Here, tf_i represents the traffic load at the base station that is temporarily deactivated, and sharing_status_i denotes the binary status of network sharing at each time point (with '1' indicating active sharing and '0' representing periods without sharing). The variable n corresponds to the total number of time points under consideration in the analysis. The table 7.1 7.2 7.3 provides averages for network sharing efficacy in different regions and between operators, as well as the density of two operators' base stations in three distinct regions.

The outcomes of this case study, particularly focusing on the suburban region, indicate a significant potential for enhanced traffic load transfer between base station pairs in a network sharing arrangement. It is observed that, within both city center and suburban areas, the traffic load transfer from Orange to Bouygues Telecom exhibits a slight increase. However, this increase does not manifest a clear correlation with the density of the operators' base station infrastructure. The geographical distribution of base stations and the average traffic load during intervals amenable to network sharing emerge as pivotal factors. These results imply that it is really complicated to understand the potential factors for network sharing performance.

Region Average

Region	Average
City Center	0.689170337
Suburban	0.749352319
Rural	0.715044377

Table 7.1: Regions' average traffic load transfer

Operator Average

Transfer	Average
City Center Orange to Bouygues	0.861828366
City Center Bouygues to Orange	0.602841322
Suburban Orange to Bouygues	0.761416507
Suburban Bouygues to Orange	0.713159754
Rural Orange to Bouygues	0.675950492
Rural Bouygues to Orange	0.73459132

Table 7.2: Operators' average traffic load transfer

The Density of Two Operators in Three Regions

Region	Orange Base Stations/km^2	Bouygues Base Stations/km^2
City Center	6.25	6.375
Suburban	2.5	2
Rural	0.875	1

Table 7.3: The Density of Two Operators in Three Regions

7.3 Percentage of network sharing f_T

Regarding the conceptual framework of network sharing, it is characterized by two distinct states, predicated on the activation status of a secondary base station. The proportion of time during which network sharing is operational is quantified by the formula:

$$f_T = \frac{\text{Number of period\_share}}{\text{Total periods number}}$$

This study approaches the analysis from two perspectives. Initially, the day is segmented into four periods: 00:00-06:00, 06:00-12:00, 12:00-18:00, and 18:00-24:00, (total 720 small time intervals for each period in one month)to assess

potential fluctuations in network sharing throughout the day. Table 7.4 suggests a noticeable decline in f_T as the day progresses, irrespective of geographic region. This trend could be attributed to the surge in social media usage and video streaming activities, particularly between 18:00 and 24:00, which dominate the traffic volume, consequently reducing the feasibility of network sharing during these hours. The evaluation of f_T across different mobile network operators, with findings presented as table 7.5 shows:

ID_Sharing	P1_NS	P2_NS	P3_NS	P4_NS
city_center	716	645.75	472.5	385.875
sub_urban	714.6	641.7	478.7	479.1
rural	717.2	676.7	525.5	435
avg f_T	0.994	0.909	0.684	0.602

Table 7.4: NS percentage for different periods in one day

These results indicate a pronounced correlation between f_T and specific time periods rather than the operators themselves or their respective regions. Across various contexts, the f_T remains consistently high, ranging from 60% to 90%, underscoring the operational viability of network sharing. This stability suggests that base stations typically operate under low-load conditions, enabling a single base station within a shared pair to support the combined traffic without exceeding its capacity limits. Therefore, network sharing appears as a sustainable and practicable strategy for green networking concepts, demonstrating its feasibility across diverse temporal and landscapes.

7.4 Power percentage saved

Building upon the power consumption model mentioned in the previous chapter, the power utilization for each time interval can be accurately evaluated. The adoption of network sharing enables the temporary deactivation of a base station (BS), thereby conserving energy that would otherwise be expended during these periods. Consequently, a revised calculation of power consumption is undertaken, incorporating the combined traffic load of the network-sharing BS pair and the operational capacity of the active BS. The efficiency of this strategy is quantified through the calculation of the power percentage saved, defined as the reduction in power usage relative to the scenario where both BSs remain operational. The formula is expressed as:

$$\text{Power percentage saved} = \frac{\text{total power of bs1 and bs2} - \text{power\_shared}}{\text{total power of bs1 and bs2}}$$

Location and Operator	**NS Percentage**
City Center Orange	0.715
City Center Bouygues	0.9425
Suburban Orange	0.812
Suburban Bouygues	0.827
Rural Orange	0.833
Rural Center Bouygues	0.786

Table 7.5: NS Percentage for different regions and operators

And the result is presented in table 7.6 indicating a positive correlation between the power percentage saved and the percentage of network sharing, demonstrating that network sharing strategies facilitate significant energy conservation, potentially exceeding 40% compared to non-sharing scenarios.

Location and Operator	Power Percentage Saved	Percentage of NS
City center Orange open	0.305	0.715
City center Bouygues open	0.427	0.9425
Suburban Orange open	0.358	0.812
Suburban Bouygues open	0.366	0.827
Rural Orange open	0.367	0.833
Rural center Bouygues open	0.339	0.786

Table 7.6: Power Percentage Saved with respect to Network Sharing Percentages.

To further quantify this relationship, a linear regression analysis was employed, aiming to evaluate the linear correlation between the percentage of network sharing $f_T(x)$ and the average power savings (y) across BS pairs in each area. After excluding outliers, the resultant linear equation shown in figure 7.3: $y = 0.48x - 0.04$ reveals a positive correlation coefficient of 0.48. This coefficient suggests that increases in f_T directly contribute to energy saving. The findings from this regression analysis confirm the existence of a linear relationship between f_T and power savings, positing that maximal network sharing could yield energy savings at approximately 44%. Such huge energy conservation underscores the potential for significant reductions in electricity costs, particularly within urban settings, confirming the environmental and economic benefits of network sharing.

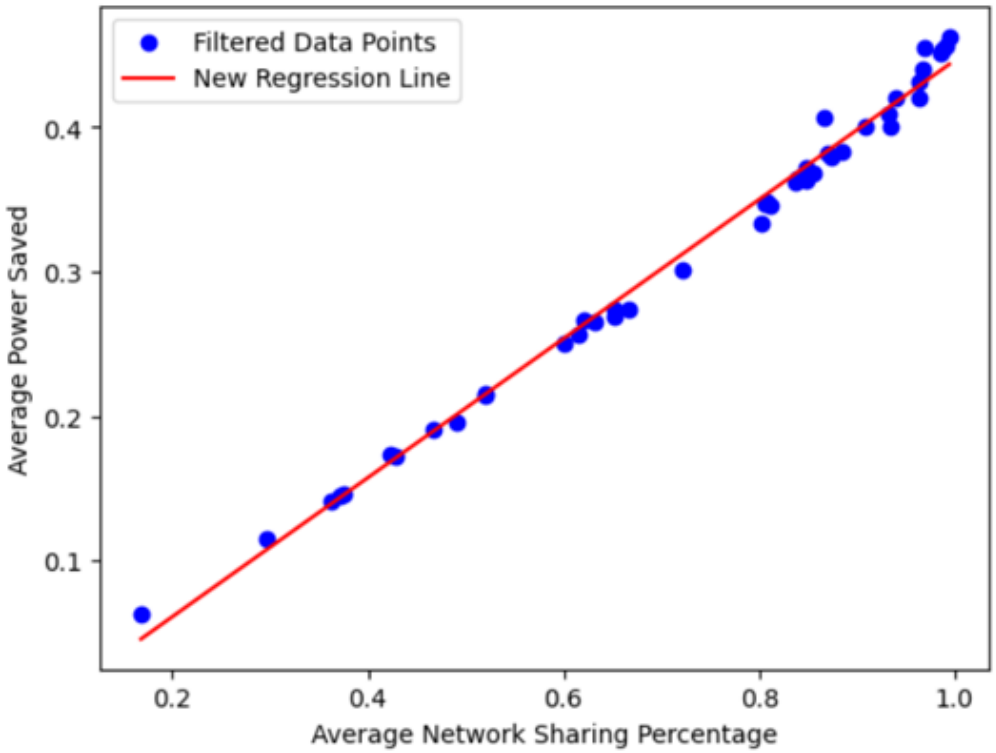

Figure 7.3: Linear regression result of the average power savings with respect to the percentage of NS

Chapter 8

Triple base station network sharing analysis

To enhance the applicability and effectiveness of network sharing (NS) strategies, this chapter explores the transition from dual to triple base station (BS) NS configurations. Notably, the covering time of NS is significantly high, this observation raises the idea of the feasibility of implementing NS strategies involving three base stations from three closely situated Mobile Network Operators (MNOs), a scenario commonly encountered in city center regions where three base stations are either co-located or positioned within a 50-meter radius of one another, as the previously established threshold. Certain BS pairs have been identified for in-depth analysis.

8.1 One threshold triple BSs NS strategy

This approach mirrors the dual strategy, wherein a single base station with the largest capacity remains active while two others are deactivated when the traffic load of three BSs falls below a pre-defined threshold Cth is set as 80% of capacity as before. Consequently, the traffic load from the two inactive stations is transferred to the operational station, thereby optimizing infrastructure utilization. The focus of this analysis encompasses two primary aspects: the comparative reduction in NS usage vis-à-vis dual BS configurations and the correlation between energy savings and NS usage. Closing two base stations obviously leads to enhanced power conservation. Figure 8.1 shows the daily trend of NS utilization. Employing triple BS NS configurations, with respect to dual configurations, results in a reduction exceeding 10% during peak traffic periods, such as lunchtime and evening, while maintaining a relatively high NS usage (over 65%) during all periods. This stability can be attributed to the design of base stations, which often possess capacities far exceeding average demands, thereby ensuring adequate residual capacity to

support the combined traffic load of three stations.

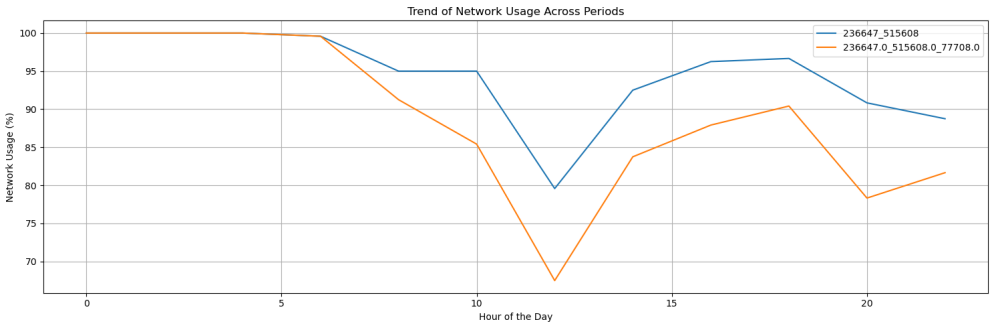

Figure 8.1: Trend of NS across the periods for double and triple NS

Data from table 8.1 comparing the NS frequency and monthly power savings (in watts) for both dual and triple BS configurations indicate that triple BS NS, despite a substantial reduction in NS usage (sometimes nearly 30%), still achieves a 5%-10% increase in monthly energy savings compared to dual-station scenarios. These findings suggest that triple BS NS could be a viable strategy when the inter-station distances among the three MNOs meet specified criteria.

Configuration	NS Usage	Power Saved Per Month
307173_double	0.788889	1209193(0.2272)
307173_triple	0.467361	1411063(0.2651)
236358_double	0.804861	1197605(0.2301)
236358_triple	0.520833	1537951(0.2955)
209633_double	0.588194	853962.8(0.1596)
209633_triple	0.395833	1175818(0.2198)

Table 8.1: Network Sharing Usage and Power Consumption Saved Per Month

To quantitatively assess this relationship, a linear regression analysis similar to previous studies was conducted to examine the linear correlation between NS frequency (x) and average power savings (y) across BS pairs in each area. The analysis, illustrated in figure 8.2, reveals a steeper slope for the triple base station scenario, indicating a more efficient power saving relative to NS frequency. However, practical implementation must consider operational constraints and the willingness of additional operators to engage in shared agreements.

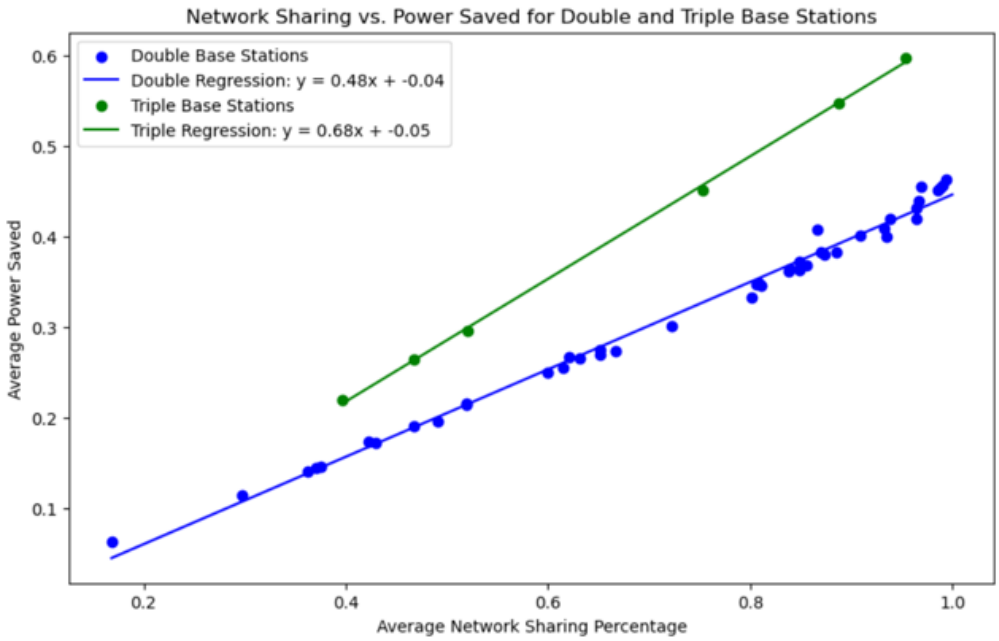

Figure 8.2: Linear function conclusion of double and triple BS NS

8.2 Dual threshold triple BSs NS strategy

Following the previous research, it was observed that the fT of triple NS appears to be relatively low. This finding suggests an opportunity to enhance the network sharing framework by introducing a dual-threshold strategy, which could potentially yield a more optimized system. The proposed two-threshold strategy functions as follows: when the traffic volume is beneath the lower threshold—defined here as the "small traffic volume" threshold—network demand can be met by maintaining a single base station, specifically the one with the highest capacity, in active mode. This approach capitalizes on periods of low traffic, maximizing energy savings while still catering to service requirements. Conversely, if the traffic volume surpasses the first threshold yet remains below a higher secondary threshold, the strategy calls for keeping the two base stations with the largest capacities operational. This intermediate step ensures that moderate increases in traffic can be accommodated without fully activating all network resources. Finally, should the traffic volume exceed the second, higher threshold, the strategy dictates that all three base stations remain operational to adequately handle periods of peak demand.

The analytical formulation for determining the operational thresholds based on the traffic load (T_a, T_b, T_c) for base stations A, B, and C, respectively, is articulated as follows:

1. If $T_a + T_b + T_c < \max(C_a + C_b + C_c) \times 0.8$, then opening a single base station is sufficient to manage the entire traffic load.
2. If $\max(C_a + C_b + C_c) \times 0.8 < T_a + T_b + T_c < \max(C_a + C_b + C_c) \times 0.8 + \text{medium}(C_a + C_b + C_c) \times 0.8$, then two base stations should be operational.
3. If $T_a + T_b + T_c > \max(C_a + C_b + C_c) \times 0.8 + \text{medium}(C_a + C_b + C_c) \times 0.8$, then all three base stations need to be open to accommodate the traffic load.

Adopting a dual-threshold strategy allows for a more efficient utilization of infrastructure, especially during periods when traffic volume falls within a moderate range. With a two-threshold system, one base station can remain inactive in the middle traffic interval, potentially yielding greater energy savings and enhancing the fT that NS is viable. When researching on this strategy, our research focuses on three problems

- **Optimal Traffic Load Distribution:** Determining the most effective method for distributing the traffic load across two active base stations during the traffic load in intermediate range to ensure NS efficiency.
- **NS Percentage Optimization:** Evaluating how the proposed dual-threshold strategy influences the overall NS percentage, particularly in comparison to a dual base station NS setup.
- **Energy Savings Correlation:** Determining the linear relationship between the energy conservation and the NS percentage, or the number of base stations in operation.

8.3 Traffic flow strategy (How to calculate ideal power for the two base station condition)

Assuming C_a is the maximum capacity among the three base stations (BSs), C_b is the second highest capacity, X represents the traffic load assigned to base station A, and T denotes the current combined traffic load of the three base stations, such that $T = T_a + T_b + T_c$, we have the following:

Note:

- BS power consumption $P_{\text{in}} = 504 + 16.8 \cdot P_{\text{out}}$ (W), where P_{out} is the relative RF output power.
- $$P_{\text{out}} = \left(\frac{\text{traffic load}}{\text{capacity of base station}} \right) \cdot P_{\max}$$

The objective function is defined as:

$$\min\left(\frac{X}{C_a}+\frac{T-X}{C_b}\right)$$

Subject to the constraints:

$$0 < X \leq C_a < T,$$
$$0 < C_b \leq C_a.$$

The derivative of the objective function with respect to X is $\frac{1}{C_a}-\frac{1}{C_b}$.

When $C_a > C_b$, the derivative is less than 0, indicating that the objective function is decreasing. Thus, the minimum occurs at the maximum value of X, which is $X = C_a$.

The objective function then simplifies to:

$$1+\frac{T-C_a}{C_b}$$

The final power calculation for two base stations becomes:

$$\left(1+\frac{T-C_a}{C_b}\right)\cdot P_{\max}\cdot 16.8+504\cdot 2$$

If $C_a = C_b$, the derivative of the objective function with respect to X is 0, indicating that the distribution of traffic load does not impact the total power consumption.

In our predefined configuration, each base station (BS) possesses a distinct capacity. Consequently, the computation of our dual-threshold strategy assumes an ideal scenario wherein the traffic load from the deactivated base station is initially offloaded to the base station with the greatest capacity. Subsequently, the residual traffic is transferred to another active base station. Employing this method of calculation allows us to estimate the power consumption under an optimal NS arrangement and to determine the corresponding energy savings. The table 8.2 presents a comparison of various BS across three distinct NS strategies. When employing a three-base station NS strategy with two thresholds, as opposed to the dual base station setup or the initial triple base station approach with a single threshold, we observe a significant enhancement in network sharing efficiency. In particular, there is a 10% improvement over the triple base station strategy with a single threshold and an approximate 15% increase in monthly energy savings over the dual base station scenario. According to these results, the dual-threshold approach for three base stations performs better in terms of energy saving and NS efficiency. This seems to be the best NS strategy, assuming that the BS infrastructure is distributed in a way that makes triple base station NS configurations easier to perform.

Configuration	**NS Usage**	**Power Saved Per Month**
307173_double	0.788889	1209193 (0.2272)
307173_triple_1threshold	0.467361	1411063 (0.2651)
307173_triple_2threshold	0.9424	2000254 (0.3759)
236358_double	0.804861	1197605 (0.2301)
236358_triple_1threshold	0.520833	1537951 (0.2955)
236358_triple_2threshold	0.9271	2036901 (0.3914)
209633_double	0.588194	853962.8 (0.1596)
209633_triple_1threshold	0.395833	1175818 (0.2198)
209633_triple_2threshold	0.8934	1703452 (0.3183)

Table 8.2: Comparison of Network Sharing Usage and Power Consumption Savings

Chapter 9

Switch operation optimization

As previous chapter mentioned, the frequency of switch operations are thought as a critical factor in evaluating the practicality of implementing a NS strategy in real-world scenarios. The underlying objective is to minimize the frequency of these operations. Several metrics will be introduced and defined to investigate more effective constraints on switch operations, with the aim of simplifying operational complexity and enhancing the feasibility of NS strategy. This research of switch operations is conducted within the dual base station scenarios.

9.1 Switch parameters

The table 9.1 displays results of the switch efficiency in the 3 areas, segmented into various periods over the span of a day. For the purposes of this analysis, each day is divided into four six-hour periods: 00:00-06:00, 06:00-12:00, 12:00-18:00, and 18:00-24:00. This division makes it easier to evaluate possible changes in the frequency and efficiency of switch operation at different times of the day.

ID_SharingPair	P1_SE	P2_SE	P3_SE	P4_SE
City Center	1221.505	272.4136	100.6467	71.38559
Suburban	852.37	167.895	74.01043	75.09637
Rural	1240.466	352.4975	106.7782	66.55757
Average	1104.78	264.2687	93.81176	71.01318

Table 9.1: Switch Efficiency Across Different Periods and Locations

Although the results cannot conclude the relationship between switching efficiency and specific geographic regions due to the limited sample of regions, there is a clear correlation found in relation to the time of day. In particular, switch efficiency shows a noticeable decrease during the day, and periods 3 (12:00-18:00) and 4 (18:00-24:00) require special attention. These intervals show a pattern of discontinuous network sharing, characterized by instances where the traffic load approaches or exceeds the network capacity, resulting the activation or deactivation of a base station (switch state operation). The observable decrease in switch efficiency during the latter half of the day underscores the critical nature of these periods, during which the traffic load frequently fluctuates around the capacity threshold. The observed trend indicates that switch operation management must be implemented strategically to effectively improve the practicality of NS.

9.2 Fix period constraints

After dividing one day into four distinct periods, it is observed that the frequency of state changes varied across each interval. A preliminary strategy to enhance switch efficiency involves ensuring that once NS is initiated, it should be sustained for a minimum duration of 30 minutes (2 time periods). This strategy, known as "forbidden fast changes," seeks to eliminate the occurrence of short-term "010" patterns, which are marked by abrupt NS status switching. Switch efficiency has been clearly increased by putting this strategy into practice when compared to previous metrics. This quick switch (only 15 minutes of NS) is not what's considered ideal for MNOs. When this strategy is applied, Table 9.2 and table 9.3 compares the results and shows that switch operations (state changes SC in table) are significantly reduced while most of the usage of NS is maintained. Building on this, the following step is establishing an upper limit for switch operations within a single period. By setting a thoughtful threshold for the number of permissible state changes:

- When detecting the fourth state change within a period, subsequent actions depend on the state of the switch. If the transition is from NS active (1) to NS inactive (0), all subsequent states within that period are set to 0. For instance, from an original sequence of '`011101110`', following the final switch to state 0, all remaining NS statuses for that period would similarly be adjusted to 0.

- Conversely, if the switch is from NS inactive (0) to NS active (1) and occurs for the fifth time within a period, all states following this fifth transition are set to 0. An illustrative sequence might transform from '`111011101111`' to '`1110111011110000`' after adjustment.

ID_pair	P3_SC	P3_NS	P3_SE	P4_SC	P4_NS	P4_SE
307189_622919	207	413	29.92754	179	289	24.21788
535932_211336	52	679	195.8654	77	655	127.5974
550114_218981	71	669	141.338	83	654	118.1928
682471_340649	46	687	224.0217	122	619	76.10656

Table 9.2: NS and SE Original

ID_pair	P3_SC	P3_NS	P3_SE	P4_SC	P4_NS	P4_SE
307189_622919	145	381	39.41379	120	261	32.625
535932_211336	48	677	211.5625	61	647	159.0984
550114_218981	63	665	158.3333	69	647	140.6522
682471_340649	40	684	256.5	106	611	86.46226

Table 9.3: NS and SE After Forbidden Fast Changes

9.3 Switch trade off parameters

To evaluate the effectiveness of different operational constraints strategies, it is important to introduce a metric that quantifies the balance between energy conservation and the reduction in switch operations. This metric, referred to as the Average Trade-off, is calculated as follows:

$$\text{Average Trade-off} = \frac{\text{energy saved reduced due to the strategy}}{\text{switch operation reduced}}$$

The objective is to minimize this metric, with the optimal scenario being a significant reduction in switch operations without a corresponding decrease in energy savings. This metric serves as a benchmark for comparing the effectiveness of two NS strategies: the fast forbidden changes strategy and a strategy that sets upper bounds on the fast forbidden changes framework. The comparative analysis of these strategies, particularly their trade-offs, is presented in table 9.4:

The analysis's conclusions show that the adopted strategy is effective, as shown by the low average trade-off seen in the city center region. Moreover, this trade-off value shows a trend of decreasing as upper bounds are established. These findings support the fixed period method as an effective strategy for improving operational viability.

Strategy	City Center	Suburban	Rural
Forbidden “010”	0.13	0.5735	0.5013
Forbidden “010” and add “upper bounds”	0.06	0.4376	0.4978

Table 9.4: Average trade off after setting Upper Bounds

9.4 Sliding window constraints

After investigating the previously mentioned switch operational strategy, a more flexible and dynamic method is discovered. It is based on the idea of a sliding window that moves across the NS state time intervals. To be more precise, if the frequency of status changes during that window exceeds a predefined threshold: all statuses until the sliding window ends are changed to a state that stops NS (referred to as the 0 state). The sliding window setting is defined by two key parameters: the length of the window and the operation threshold. A figure 9.1 is presented to clarify the effects of these parameters, particularly state changes caused by reaching the operational threshold (known as cooling state changes). Cooling state changes occur when the original active NS period (indicated by 1) is forced to become inactive (indicated by 0) because of the sliding window cooling strategy.

The figure quantifies the impact of this strategy on NS time points, indicating how many of them are affected. The analysis reveals that larger window sizes are more sensitive to detecting changes that trigger the cooling strategy, because they pass more data points. For each window size, as the threshold increases, the average cooling changes generally decrease. Conversely, smaller windows and lower thresholds demonstrate greater reactivity, adjusting more frequently to immediate changes, whereas larger windows and higher thresholds adopt a more conservative stance. The optimal window size and threshold combination is determined by the cooling strategy's specific objectives. For example, if the goal is to value significant changes, using larger window sizes and higher thresholds may be beneficial. This strategic flexibility enables adapted adjustments to NS operations, increasing network management efficiency and adaptability. To assess the effectiveness of this strategies, the same metric parameters used in the fixed period strategy were used for comparison analysis. Within a six-hour period, with a time window size of 24, up to four state changes are permitted.

9.5 Comparision results

The results from table 9.5 enables a comparative evaluation. Both the sliding window and fixed period strategies resulted in an increase in switch efficiency over

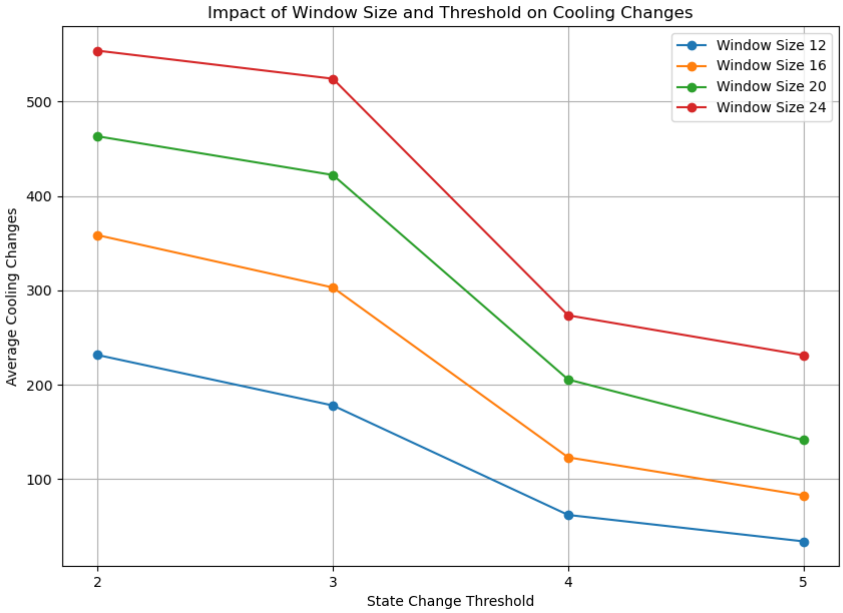

Figure 9.1: Impact of window size and threshold on cooling strategy

the initial conditions. This shows that each switch operation contributes to a longer period of network sharing. However, switch efficiency alone is insufficient to determine whether one strategy is better than another. This is mainly because reducing switch operations tends to result in a significant decrease in NS duration.

For instance, specific base station pairs, such as those with IDs 307189 and 622919, in sliding window scenario experienced a reduction in NS percentage to between 1/4 and 1/2 of the original levels. Conversely, the fixed period strategy typically resulted in NS percentages being maintained at approximately 3/4 of their initial values. The sliding window strategy has the potential to significantly reduce the NS percentage and frequency of switch operations. Therefore, this approach would have a larger energy saving trade offs, resulting in the calculation of Switch Trade-off parameters as shown in table 9.5:

The fixed period strategy appears to be wiser for several reasons. First, it has a lower energy trade-off, retaining the majority of the energy savings while limiting switch operations. Furthermore, fixed period strategies provide a consistent operational schedule, it is generally simpler, reducing the complexity in adjusting to real-time data. When setting the upper bound of the switch, the sliding window strategy may have a greater tradeoff for energy savings. That could be because the sliding window approach is sometimes too sensitive to short-lived fluctuations,

ID_SharingPair	Sliding Window	Fix Period
307189_622919	0.034501	0.015308
535932_211336	0.046847	0.021014
550114_218981	0.057317	0.019444
682471_340649	0.053495	0.020833

Table 9.5: Average Trade-off Comparison of Sliding Window and Fix Period

resulting in the loss of more potential sharing opportunities. (For example, if there is a fluctuation at 23:00, the method will give up the network sharing opportunities at midnight, but the fix strategy method will not have this concern because the threshold refreshes at 24:00.)

Chapter 10

Prediction of traffic load and status of NS

Predicting traffic loads at BSs is an important aspect of this research, especially for prepared operational adjustments rather than real-time reactions. The essence of 'advance' adjustments implies the need to forecast future data points within the traffic series using historical data. This predictive capability allows for the determination of whether network sharing (NS) can be activated at upcoming time intervals, thereby increasing the practicability of the NS strategy. For such predictive analysis, various machine learning methods for forecasting the combined traffic load of base station pairs may be used. This preliminary step is critical for determining the feasibility of implementing the NS strategy in the future. By accurately predicting traffic load, the system can adjust NS parameters ahead of time, ensuring peak network performance and resource utilization.

10.1 Base station traffic load prediction with LSTMs

10.1.1 Prediction with small training size

By employing the Long Short-Term Memory (LSTMs) methodology, the initial attempt utilizes data from a singular pair of combined base stations for prediction purposes, with the parameters configured as follows:

- `train_size = 672`, means that we use one week's data as the training set, one week's data could contain a complete trend of traffic trace.
- `test_size = 96`, the testing set is one day to see whether the trend fits the origin.

- `look_back = 24` means that use every 6 hours' data to predict the subsequent time period.

Hyperparameters:

- `epochs = 50`,
- `batch_size = 16`,
- `optimizer = Adam`,
- `learning_rate = 0.001`.

It is important to note that *look back* refers to the number of prior time intervals utilized as input variables to predict the subsequent time period in a time series. The findings of this investigation reveal a Test RMSE of 0.117, a Test R^2 of 0.638, and a Test MAPE of 50.212%. These results, illustrated in figure 10.1 and detailed above, indicate a less than ideal prediction performance, with an R^2 value of only 0.638 suggesting the model's limited accuracy and its inability to capture more than the general trend.

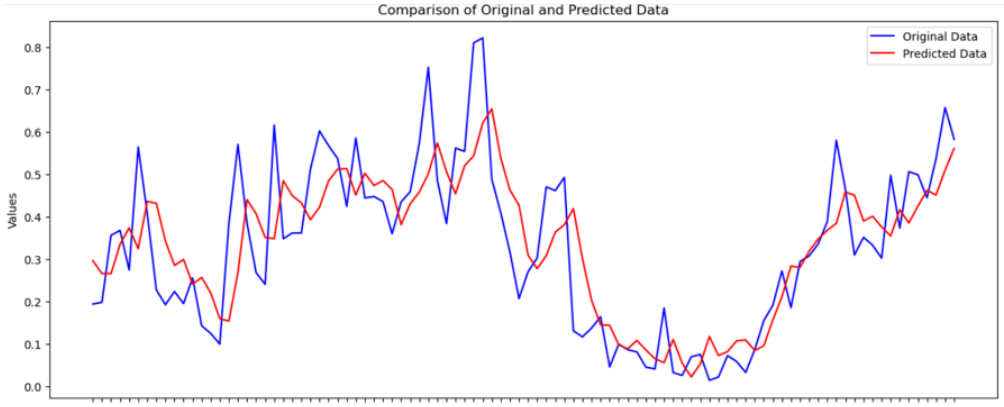

Figure 10.1: LSTMs traffic load prediction with the smaller size of training set

10.1.2 Prediction with larger training size

The training dataset's insufficient size is likely the reason for the initial lack of prediction accuracy, as it does not fully capture the entire periodic characteristics of the data. Following optimizations, the training dataset was increased to include information from four pairs of combined base stations (one pair was used as the test set), all of which were located in city centers. The updated configurations are:

- `test_size = 2832` one month data of one pair BS has 2880 time periods and total number of time periods – look_back= test_size
- `look_back = 48`

Hyperparameter

- `epochs=50`, `batch_size=16`
- `optimizer = Adam`

The results: RMSE: 0.0860 R^2: 0.816
Figure 10.2 illustrates the significant improvement in the improved predictions. However, a persistent challenge is the model's failure to accurately predict peak traffic volumes. This limitation indicates a potential bias in the model toward lower traffic scenarios, which might have arisen from the proportion of low-traffic cases represented in the training data during the middle of the night. The inability to accurately predict the peak value may influence the determination of whether NS can be used. An analysis of predicted versus original NS usage, segmented into twelve two-hour periods throughout a day. The analysis presented in figure 10.3 reveals a tendency to overestimate network sharing (NS) usage across the evaluated time points. This overestimation is particularly pronounced during Periods 6 and 9, corresponding to the intervals of 12:00-14:00 and 18:00-20:00, respectively. These periods are critical as the traffic load approaches the operational threshold Cth, effecting the accuracy of predictions. Minor deviations in forecasting can lead to incorrect predictions regarding the activation of NS, underscoring the challenges in predicting traffic loads near threshold levels.

The analysis of 2832 time points reveals some inconsistent points in NS status predictions, with a total of 423 instances where the predicted status differed from the original status—either predicting NS status is 1 or vice versa. Despite this, the predictive model identified 2371 time points as suitable for NS, compared to the actual data of 2228, yielding an 85.1% accuracy. This performance metric demonstrates the machine learning (ML) approach's ability to accurately forecast NS applicability. The ML method is practical and beneficial for simplifying operator operations and real-time monitoring steps, completing policy deployment ahead of time, and achieving the goal of sustainable green networks through shared infrastructure.

10.1.3 Prediction with second half day's data

An experiment was carried out focusing on the prediction of the combined traffic load of base stations using only data from the latter half of the day to address potential biases towards lower traffic loads that exist in the data. This approach

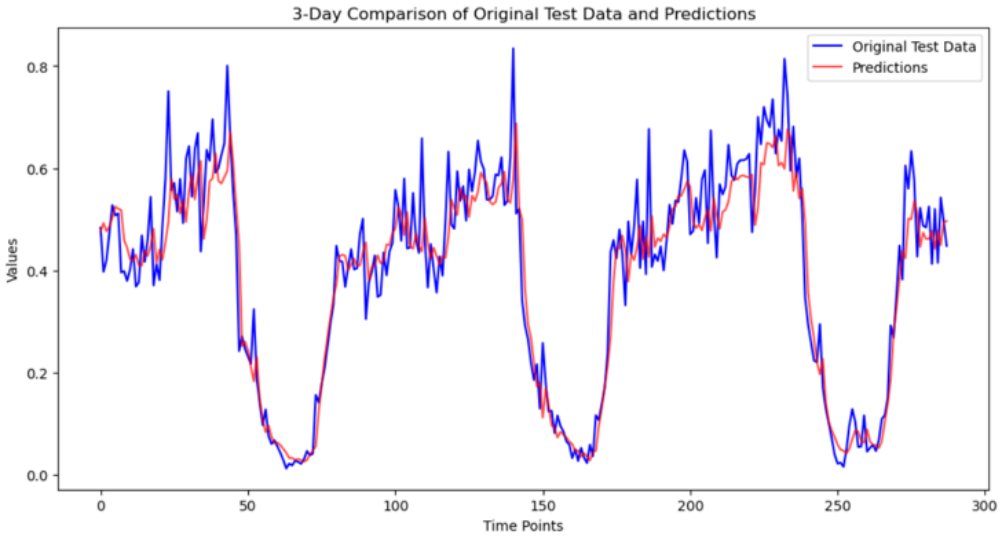

Figure 10.2: LSTMs traffic load prediction with larger size of training set

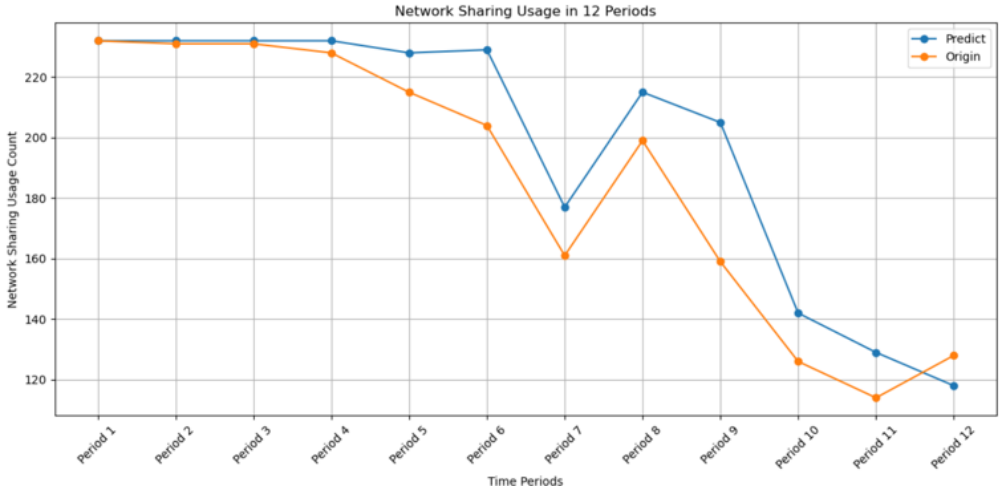

Figure 10.3: NS usage prediction and original

was predicated on the assumption that the variability and peak loads characteristic of afternoon and evening traffic (12:00 to 24:00) could provide a more challenging dataset for predictive analysis.

For the analysis, traffic load data from the afternoon and evening over one month were aggregated and then subjected to max-min normalization. This preprocessing step ensures that the data values are scaled within a specific range, facilitating more efficient learning by the LSTM model. The experiment was structured with four pairs of combined base station data serving as the training set and one pair as

the test set, all selected from base stations located within city center areas.

Hyperparameter settings for the experiment were as follows:

- Epochs: 30,
- Batch size: 16,
- Optimizer: Adam,
- Learning rate: 0.001,
- Look back: 96.

Upon completion of the training process, the model yielded an RMSE of 0.1226 and a Test R^2 of 0.3526. The notably low R^2 value underscores the model's difficulty in accurately predicting traffic loads during these periods, attributed to the afternoon and evening hours' notable fluctuation and peak traffic volumes. The model's overall predictive accuracy was clearly impacted by excluding morning data from the analysis, which is typically more uniform and predictable. This highlights the difficulty of forecasting in the presence of notable fluctuations and peaks in traffic load.

10.1.4 Prediction of network sharing status

The subsequent phase explores the prediction of network sharing (NS) status, focusing specifically on binary outcomes (1 meaning NS is allowed, 0 meaning NS is not allowed). The goal is to predict future states by utilizing past NS utilization patterns. But it's important to recognize that every base station (BS) pair displays different NS usage patterns. For this reason, the analysis employs data from a singular pair of combined base stations, with 70% of the data allocated for training and the remaining 30% for testing. The `look_back` parameter is set to 96.

The predictive analysis utilized Random Forest and Long Short-Term Memory (LSTMs) models, with the outcomes presented as follows:

- Logistic Regression Accuracy: 0.7667464114832536
  Classification Report:
  - For status 0: Precision of 0.52, Recall of 0.73, and F1-Score of 0.61.
  - For status 1: Precision of 0.90, Recall of 0.78, and F1-Score of 0.83.
- Method: LSTMs Accuracy: 0.7168458781362007
  Classification Report:
  - For status 0: Precision of 0.74, Recall of 0.60, and F1-Score of 0.66.

– For status 1: Precision of 0.71, Recall of 0.81, and F1-Score of 0.76.

Even though it achieved a remarkable accuracy rate of approximately 0.767, using only the binary status to predict NS ability might not be the most effective approach. This results from the model's learning from the binary status pattern, which ignores the critical role of real traffic load, which is the main variable influencing the NS status. For this reason, there's a chance that the model will become overfit to past patterns of status changes and miss the real dynamics guiding these changes.

10.2 Base station traffic load prediction with other ML method

In addition to exploring Long Short-Term Memory (LSTMs) models, this research extends its analysis to incorporate other machine learning (ML) methods for predicting base station traffic loads. Specifically, we explore the applications of Extreme Gradient Boosting (XGBoost) and Autoregressive Integrated Moving Average (ARIMA) methodologies.

10.2.1 Base station traffic load prediction with XGBoosts

The XGBoost model was applied using data from three pairs of combined base stations as the training set and data from an additional pair for testing, all selected from city center locations. The configuration for this model was as follows:

- Prediction size was set to 2784, indicating the total number of predictions the model was expected to make.
- WINDOW_SIZE, established at 96, signifies the quantity of consecutive data points employed to generate a single model input. This parameter illustrates the implementation of a sliding window technique for feature generation, where the previous 96 observations are utilized to forecast the subsequent data point.
- PREDICT_AHEAD is designated as 1, specifying that the model's forecast horizon spans one future time step. This suggests an emphasis on short-term prediction, aiming to estimate the traffic load at the next immediate point by analyzing the data within the preceding WINDOW_SIZE interval.

The obtained results were as follows:

- MAPE (Mean Absolute Percentage Error): 19.4488%,
- R^2 (Coefficient of Determination): 0.8021,

- Accuracy: 84.1%.

The predictive performance depicted in the resulting figure 10.4 shows good overall curve alignment with actual data trends. However, it also reveals a limitation in accurately capturing peak traffic fluctuations, implying that while XGBoost effectively models the general traffic pattern, it challenges to accurately predict peak load variations.

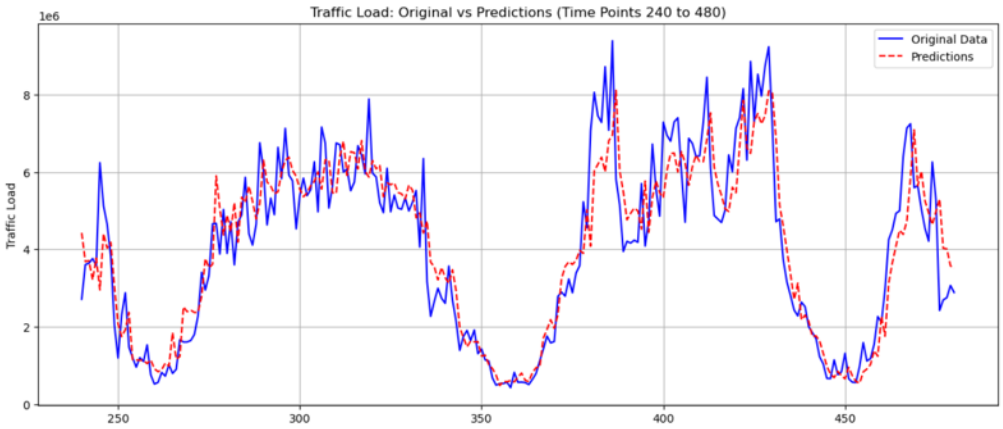

Figure 10.4: XGBoost traffic load prediction results

10.2.2 Base station traffic load prediction with ARIMA

In furtherance of the study on base station traffic load prediction, an exploration was conducted using the Autoregressive Integrated Moving Average (ARIMA) model, focusing on a single pair of combined base station traffic data. The analysis was structured around specific configurations:

The ARIMA model's hyperparameters were carefully selected as follows:

1. p (Autoregressive Order): Set to 1, this parameter quantifies the extent to which the model accounts for the immediate past value in its predictions, essentially capturing the series' momentum or trend.

2. d (Differencing Order): Also set to 1, indicating a single differencing step to achieve stationarity within the series, thereby mitigating any potential variability in the overall trend over time.

3. q (Moving Average Order): Established at 1, it incorporates the forecast error from the preceding time step into the model, aiding in the refinement of future predictions.

The $WINDOW\_SIZE = 96$ and $STEP\_SIZE = 1$ settings are related to the structure of the dataset being used for the ARIMA model rather than the model itself:

- $WINDOW\_SIZE$ is commonly used to denote the number of historical data points used to make a prediction. In this case, the last 96 observations are used, which could mean the last 96 time intervals given the frequency of data collection.

- $STEP\_SIZE$ usually refers to the increment at which the sliding window moves forward across the time series. A $STEP\_SIZE$ of 1 implies that after each prediction, the window moves one interval forward.

The outcomes of the ARIMA model were quantified as follows:

- Mean Absolute Percentage Error (MAPE): 22.372%,

- Coefficient of Determination (R^2): 0.7617,

- Accuracy: 84.1%.

The ARIMA model performs well in forecasting peak traffic loads, as shown in the figure 10.5. Regardless of the inherent challenge in predicting these fluctuations, the ARIMA model's performance demonstrates a strong ability to model traffic trends effectively, pointing out its practicality in accurately forecasting peak demand periods.

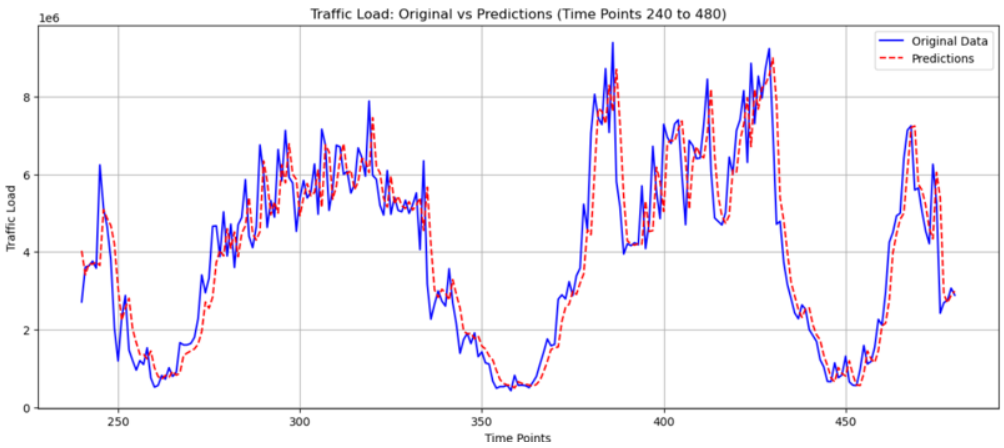

Figure 10.5: ARIMA traffic load prediction

10.3 Base station network sharing prediction

The precision of our predictive models is assessed using the formula:

$$\text{Accuracy} = \frac{\text{total predicted number of points} - \text{mismatch of status of NS}}{\text{total predicted number of points}}.$$

In predicting traffic load and network sharing (NS) status, a critical metric is false positives, which occur when time points that were not originally suitable for NS (status 0) are incorrectly forecasted to be suitable (status 1). Overestimations may exceed the predetermined threshold (C_{th}), making the base station (BS) unable to manage the high traffic volume, affecting quality of service (QoS) and user satisfaction.False negatives, where NS is feasible but predictions indicate otherwise, simply represent a missed opportunity for energy savings without risking BS capacity or quality of service.

An analysis was conducted to quantify the traffic load at points of false negatives and determine their excess over the C_{th} set in our NS strategy, which is traditionally set at 80%. If the traffic load during false negative occurrences falls within the 80%-100% range, it suggests that while traffic slightly exceeds the C_{th}, it does not exceed the BS's capacity, making NS viable without risk of overload. These are known as 'acceptable prediction results' or 'fake false positives.'

The actual issue is 'real false positives,' which occur when traffic load predictions exceed 100% of the BS's capacity, making such predictions unacceptable. According to the figure 10.6, traffic load at false positive points is mainly between 80% and 90% capacity. Only about 15% of these, or about 40 out of 2750 monthly time points, exceed the BS's capacity. This finding highlights the accuracy of ML methods in forecasting NS applicability, which improves the operational feasibility and adaptability of NS strategies.

10.4 Time series prediction method comparison

The evaluation of three advanced forecasting methodologies— LSTMs, XGBoost, and ARIMA—provides important insights into their applicability and performance in predicting network sharing viability using traffic load data. Each model achieves noteworthy accuracy from table 10.1, exceeding 80%, in predicting network sharing status, demonstrating the utility and efficiency of using time series prediction methods for operational decision-making in network management frameworks. Notably, XGBoost comes as a strong general predictor across the dataset, with the lowest Mean Absolute Percentage Error (MAPE), demonstrating its predictive capabilities.

The metrics for each method are close, but we are more concerned with peak traffic load prediction. The ARIMA model, while achieving the lowest MAPE or

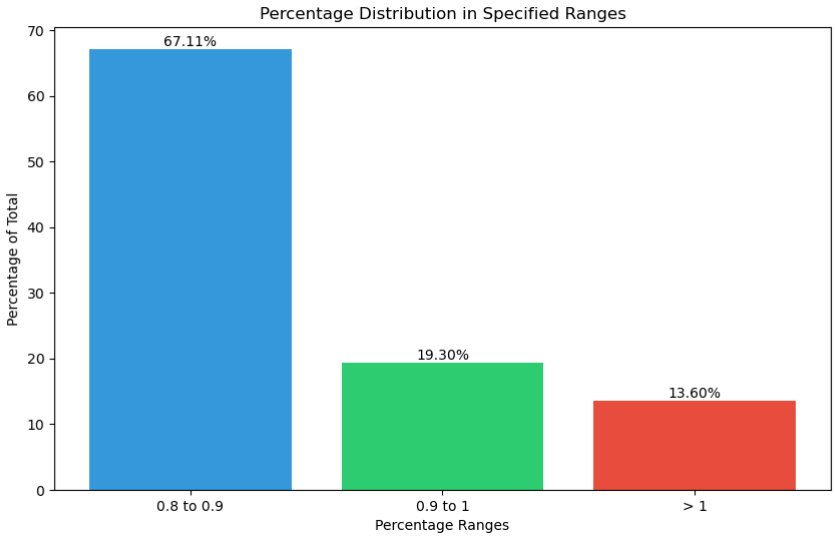

Figure 10.6: Distribution of ARIMA false positive prediction corresponds to the real percentage of BS capacity

Model	MAPE	R^2	Prediction Accuracy	FP	Real FP
LSTMs	20.356%	0.816	85.1%	283	45
XGBoost	19.4488%	0.8021	84.1%	288	41
ARIMA	22.372%	0.7617	84.1%	283	31

Table 10.1: Performance Metrics comparison

R^2, has the least false and real false positive, and offers significant advantages in peak demand forecasting. This means that ARIMA is slightly better for scenarios where accurate identification of peak periods is required. It also offers ease of use and understanding, particularly for precise peak forecasting.

Chapter 11

Conclusion

This thesis has made significant contributions to the understanding and strategic planning of mobile network infrastructure, with a particular emphasis on the analysis of base station traffic traces and their geographical distribution, which will influence the NS potential. It aimed to improve network planning and infrastructure sharing in scenarios with both double and triple base station pairs. In the context of double base station (BS) Network Sharing, the findings indicate that NS is feasible for 60-80% of the time, yielding power savings in the range of 25% to 35%. A derived linear regression model, represented by the equation power saving = 0.48 * NS percentage - 0.04, provides a quantitative basis for estimating the impact of NS utilization on energy conservation. The triple BSs NS with two thresholds improves the NS percentage to 90% and increases power savings by 15% compared to the same BS with double NS. The resultant substantial energy savings have the potential to significantly reduce electrical costs, contributing towards broader energy conservation efforts. And then the approximate result by predefined fixed period strategies that define appropriate thresholds for switch operations, this study has enhanced the practicality and efficiency of Network Sharing (NS) strategies.

Furthermore, the investigation has utilized time series machine learning techniques, including LSTM, XGBoost, and ARIMA, to forecast traffic loads and subsequent network sharing utilization. Among these methods, ARIMA showed marginally superior performance in predicting peak traffic loads. The predictive accuracy for NS utilization was approximately 85%, and after adjusting 'fake false positives', the accuracy potentially increases to about 93%. These predictive outcomes enable proactive network management, ensuring the consistent delivery of high-quality services to users.

The findings from this research offer valuable insights into the optimization of network infrastructure and underscore the viability of infrastructure sharing as a sustainable and efficient approach for the future of mobile networks. This thesis not only addresses key challenges in mobile network management but also lays a

foundation for more sustainable, and efficient telecommunications infrastructures. The strategies and methodologies proposed herein promise to revolutionize network sharing practices.

11.1 Future work

Future research directions involve expanding the network sharing strategies across larger geographic regions to gain more comprehensive insights. Graph theory may reveal more efficient methodologies for allocating traffic loads across base stations, increasing the granularity and precision of traffic trace analysis. Furthermore, classifying base stations into distinct clusters based on traffic load characteristics allows for the adaptation of NS strategies to each cluster's specific needs. This approach aims to improve NS planning by transitioning it to a more universally applicable framework that can accommodate the diverse operational dynamics of mobile networks.

Bibliography

[1] Yi Wang, Qixin Chen, Ning Zhang, Cheng Feng, Fei Teng, MY Sun, and CQ Kang. «Fusion of the 5G communication and the ubiquitous electric internet of things: application analysis and research prospects». In: *Power system technology* 43.5 (2019), pp. 1575–1585 (cit. on p. 1).

[2] Huawei. *5G Power: Green Grid Slashes Costs, Emissions, Energy Use*. https://www.huawei.com/en/huaweitech/publication/89/5g-power-green-grid-slashes-costs-emissions-energy-use. Accessed: 2023-03-12. 2023 (cit. on p. 1).

[3] Yongjun Xu, Guan Gui, Haris Gacanin, and Fumiyuki Adachi. «A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges». In: *IEEE Communications Surveys Tutorials* 23.2 (2021), pp. 668–695. DOI: 10.1109/COMST.2021.3059896 (cit. on p. 3).

[4] Vu Khanh Quy, Abdellah Chehri, Nguyen Minh Quy, Nguyen Dinh Han, and Nguyen Tien Ban. «Innovative Trends in the 6G Era: A Comprehensive Survey of Architecture, Applications, Technologies, and Challenges». In: *IEEE Access* 11 (2023), pp. 39824–39844. DOI: 10.1109/ACCESS.2023.3269297 (cit. on p. 3).

[5] Nina Slamnik-Kriještorac, Haris Kremo, Marco Ruffini, and Johann M. Marquez-Barja. «Sharing Distributed and Heterogeneous Resources toward End-to-End 5G Networks: A Comprehensive Survey and a Taxonomy». In: *IEEE Communications Surveys Tutorials* 22.3 (2020), pp. 1592–1628. DOI: 10.1109/COMST.2020.3003818 (cit. on p. 3).

[6] Fatma Ezzahra Salem, Tijani Chahed, Zwi Altman, and Azeddine Gati. «Traffic-aware Advanced Sleep Modes management in 5G networks». In: *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. 2019, pp. 1–6. DOI: 10.1109/WCNC.2019.8886051 (cit. on p. 3).

[7] Daniela Renga, Zunera Umar, and Michela Meo. «Trading Off Delay and Energy Saving Through Advanced Sleep Modes in 5G RANs». In: *IEEE Transactions on Wireless Communications* 22.11 (2023), pp. 7172–7184. DOI: 10.1109/TWC.2023.3248291 (cit. on p. 3).

[8] Bingyu Xu, Dantong Liu, Yue Chen, and Jesus Requena Carrion. «User association in energy cooperation enabled HetNets with renewable energy powered base stations». In: *2015 10th International Conference on Communications and Networking in China (ChinaCom)*. 2015, pp. 801–806. DOI: 10.1109/CHINACOM.2015.7498047 (cit. on p. 3).

[9] Yosof M. Khalifa, Sabria M. Syour, and Farag S. Alargt. «Optimal Design of a Hybrid Renewable Energy System Powering Mobile Radio Base Station in Libya». In: *2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*. 2021, pp. 423–429. DOI: 10.1109/MI-STA52233.2021.9464498 (cit. on p. 3).

[10] André Gomes, Jacek Kibiłda, Arman Farhang, Ronan Farrell, and Luiz A. DaSilva. «Multi-Operator Connectivity Sharing for Reliable Networks: A Data-Driven Risk Analysis». In: *IEEE Transactions on Network and Service Management* 18.3 (2021), pp. 2800–2811. DOI: 10.1109/TNSM.2021.3073841 (cit. on p. 3).

[11] M. Ajmone Marsan, M. Meo, M. Ni, and D. Renga. *Network Sharing for sustainability and resilience in the era of 5G and beyond*. NetMob 2023. URL: https://netmob.org/book%5C_of%5C_abstract/book%5C_of%5C_abstract.pdf (cit. on p. 3).

[12] Marco Ajmone Marsan and Michela Meo. «Network sharing and its energy benefits: A study of European mobile network operators». In: *2013 IEEE Global Communications Conference (GLOBECOM)*. 2013, pp. 2561–2567. DOI: 10.1109/GLOCOM.2013.6831460 (cit. on p. 3).

[13] Qi Wang and Jun Zheng. «A Distributed base station On/Off Control Mechanism for energy efficiency of small cell networks». In: *2015 IEEE International Conference on Communications (ICC)*. 2015, pp. 3317–3322. DOI: 10.1109/ICC.2015.7248836 (cit. on p. 3).

[14] Xiaoxing Yu and Jing Feng. «Research of base station spatial distribution model based on real metropolis data». In: *2016 IEEE Region 10 Conference (TENCON)*. 2016, pp. 2090–2093. DOI: 10.1109/TENCON.2016.7848394 (cit. on p. 4).

[15] Sheng Zhou, Dongheon Lee, Bingjie Leng, Xuan Zhou, Honggang Zhang, and Zhisheng Niu. «On the Spatial Distribution of Base Stations and Its Relation to the Traffic Density in Cellular Networks». In: *IEEE Access* 3 (2015), pp. 998–1010. DOI: 10.1109/ACCESS.2015.2452576 (cit. on p. 4).

[16] Basma Mahdy, Hazem Abbas, Hossam Hassanein, Aboelmagd Noureldin, and Hatem Abou-zeid. «A Clustering-Driven Approach to Predict the Traffic Load of Mobile Networks for the Analysis of Base Stations Deployment». In: *Journal of Sensor and Actuator Networks* 9 (Nov. 2020), p. 53. DOI: 10.3390/jsan9040053 (cit. on p. 4).

[17] Rosario G. Garroppo and Christian Callegari. «Prediction of mobile networks traffic: enhancement of the NMLS technique». In: *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. 2020, pp. 1–6. DOI: 10.1109/CAMAD50429.2020.9209314 (cit. on p. 4).

[18] Orlando E Martínez-Durive, Sachit Mishra, Cezary Ziemlicki, Stefania Rubrichi, Zbigniew Smoreda, and Marco Fiore. *The NetMob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography*. arXiv:2305.06933 [cs.NI], 2023. arXiv: 2305.06933 [cs.NI] (cit. on p. 9).

[19] *Agence Nationale des Fréquences (ANFR)*. URL: https://data.anfr.fr,%20https://cartoradio.fr%20[Accessed%20on%2028%20June%202023] (cit. on p. 10).

[20] Gunther Auer et al. «D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown». In: *Earth* 20.10 (2010) (cit. on p. 15).