ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

TELECOMMUNICATION ENGINEERING GRADUATE PROGRAM

Design Optimization of Tracking Area in LTE Networks Using Clustering Techniques

(THE CASE OF ADDIS ABABA, ETHIOPIA)

By

Zelalem Tesfaye

Advisor

Dr. –Ing. Dereje Hailemariam

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Conditions for the Degree of Masters of Science in Telecommunication Network Engineering

December 2021

Addis Ababa, Ethiopia

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

TELECOMMUNICATION ENGINEERING GRADUATE PROGRAM

Design Optimization of Tracking Area in LTE Networks Using Clustering Techniques

(THE CASE OF ADDIS ABABA, ETHIOPIA)

By

Zelalem Tesfaye

Approved by Board of Examiners

Dr. Bisrat Derebssa	
(Chairman, School Graduate Committee)	Signature
Dr. –Ing. Dereje Hailemariam	
Advisor	Signature
Dr. Yalemzewd Negash	
Internal Examiner	Signature
Dr. Sosina Mengistu	
Internal Examiner	Signature

Declaration

I declare that this thesis is compiled and written exclusively by me. Except as noted in the text, everything of the work contained within is my own original work and no one submitted this work for any other degree or professional certification.

Zelalem Tesfaye

__________________________ __________________________

Signature

Advisor

__________________________ __________________________

Dr. –Eng. Dereje Hailemariam Signature

December, 2021

Addis Ababa, Ethiopia

Abstract

The intense competition in the telecommunications market forces operators and service providers to require optimized and cost-effective networks. In the study of Long Term Evolution (LTE) network location management, one of the major challenges is tracing users at a low cost. In LTE networks, a Tracking Area (TA) is a logical group of cells that are used to represent the location of a User Equipment (UE) for paging. The location management signaling overhead of Tracking Area Update (TAU) versus paging is a documented performance concern in LTE networks. To respond to dynamic changes in UE distribution and UE mobility patterns, TA design must be modified over time, because the initial deployment may not be efficient in terms of signaling overhead. This thesis is focused on the planning and optimization of tracking area configurations by taking a real LTE network topology in the city of Addis Ababa, Ethiopia. One of the issues addressed in the thesis is the investigation of a baseline TA configuration for overall signaling overhead by transforming UE traces in the Call Detail Records (CDR) onto the drivable road network of the city and then rebuilding their corresponding UE traces. Two heuristic planning and optimization approaches are implemented to assign and reassign cells to new TAs to find out TAs that improve the overall signaling overhead in the network studied.

The baseline configuration has 650 sites grouped into 8 TAs having an overall signaling overhead of 6.08E+08 messages. The first heuristic algorithm, which embeds the K-means clustering approach, has created 74 TAs with a minimum overall signaling overhead of 3.07E+08 messages, which is a 49.51 % improvement in total signaling overhead as compared to the baseline configuration. The total signaling overhead improvement achieved by the second heuristic algorithm using the Hierarchical Density-Based Spatial Clustering of Application with Noises (HDBSCAN) clustering approach, which created 237 TAs with a minimum overall signaling overhead of 2.46E+08, is found to be 59.54 %.

Keywords: Signaling Overhead Optimization, MME, TA, TAL, LTE, TAU, HDBSCAN

Acknowledgment

I am thankful to my advisor Dr. –Ing. Dereje Hailemariam helped me to reach this level and provided all the support I needed in all situations. He also gave me the chance to use all my energy and welcomed my request for advice, given his tight schedule. Without his unmatched support, this research would not have been successful.

Contents

List of Figures

List of Tables

Abbreviations

2G	Second Generation Networks
3G	Third Generation Networks
CDR	Call Detail Records
CN	Core Network
CRM	Customer Relation Management
EPC	Enhanced Packet Core
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Application with Noises
HSS	Home Subscriber Service
LA	Location Area
LTE	Long Term Evolution
MME	Mobility Management Entity
PCRF	Policy and Charging Rules Function
P-GW	Packet Data Network Gateway
RA	Routing Area
S-GW	Serving Gateway
TA	Tracking Area
TAC	Tracking Area Code
TAL	Tracking Area List
TAU	Tracking Area Update
UE	User Equipment

Chapter 1

Introduction

1.1. Motivation and Background

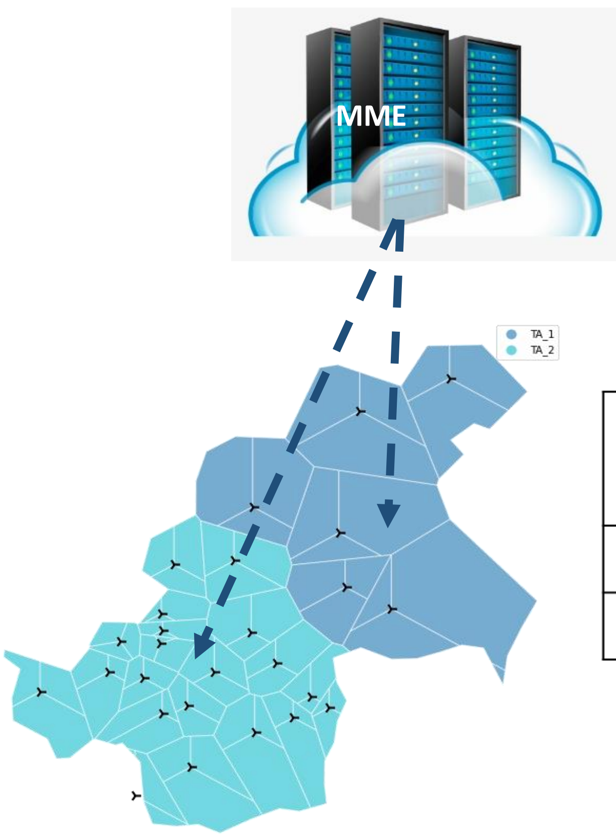
In the last few decades, the field of wireless and mobile communications has seen marvelous expansion in both capacity and coverage. Nearly 95 % of the world's population is covered with the 3rd Generation Partnership Project (3GPP) cellular technologies, one of them is Long-Term Evolution (LTE) having a projected 4.7 billion subscriptions by the end of 2021. There are more than 5.5 billion new smartphone subscriptions supporting LTE communication [1]. Despite the rapid growth, LTE networks continue to be challenged by severe signaling traffic that affects their bandwidth due to users' mobility pattern change.

To deliver service to large numbers of LTE subscribers with LTE User Equipment (UE) units, LTE networks need to know the exact location of the serving cell of each UE unit. In this circumstance, LTE networks must maintain track of the location of all UE units, and due to the high mobility of UEs, locating each UE's specific location (i.e., each UE's serving cell) within the network coverage area is difficult. The signaling load associated with tracking UE's is extremely high and amounts to 34 % of the total signaling load on the mobility management entity (MME) of the LTE network [2]. One of the primary functions of LTE networks is mobility management (MM). The network element that controls and manages an individual's UE's mobility within the LTE network is known as the MME and is tasked to handle all mobility-related messages between the UEs and the serving network [3].

There are two activity modes of UE's in LTE; namely, connected mode and idle mode [3]. The MME identifies the location of UE's at the serving cell level when the UE is in connected mode, but in the idle mode, the location of the UE's is known to the MME at the tracking area (TA) level of the LTE network. TA is a virtual region that groups many cells. The MME initiates a paging process to locate a specific idle user within the network. To put it another way, the MME sends paging signals to discover the UE's particular serving cell inside the LTE network. Tracking area update (TAU), on the other hand, is a procedure that the UE sends to the MME to tell it of a change in TA of the UE’s current location. When the mobile user is in idle mode, TA is utilized to reduce the signals created by the location management functions (i.e. paging and TAU). The MME usually finds the mobile user based on the UE's most recent location, which means the MME possesses the UE's most recent TA. The MME pages all of the cells in that TA till it receives confirmation from the mobile UE’s serving cell. By means of TAU procedure, the user updates the MME every time it moves from one cell to another that is not in the same TA of the LTE network.

However, the TA concept has several weaknesses, such as the service interruption that occurs when a cell is allocated to a different TA, as the eNodeB needs restarting for the change to take effect [4], [5]. Moreover, there is a ping-pong consequence created by edge users who are situated between two neighbor TAs [6]. 3GPP presented a new idea known as tracking area list (TAL) as an upgrade to the TA concept. TAL is a collection of TAs to provide a more flexible and diverse set of cells inside it. TAL, which works in the same way as TA, helps to reduce the signaling overhead caused by TAU. TAU procedure is triggered when a user moves from one cell to another cell that is outside of the TAL assigned to the user’s UE. The MME, in contrast, pages the user by the last known TAL to which it was registered in the network. TAL has several advantages over traditional TA, particularly in terms of reducing total signaling overhead incurred by paging and TAU. This is due to the TAL's flexibility and ability to accommodate a variety of TAs. As a result, the ping-pong consequence at TA boarders is reduced significantly [7].

In terms of signaling, the size of each TAL has a direct impact. To put it another way, when the number of cells inside a TAL grows, the paging signaling load grows while the TAU signaling load reduces, and vice versa. The standard TAU and paging systems are to blame for this as TAU happens a user moves out of a TA, while paging is broadcasted to all of the cells in the TA, simultaneously [8]. Figure 1.1 depicts how cells and TAs are grouped to generate TALs for a real deployment in the city of Addis Ababa. TA_1 is made up of 18 cells configured within 6 eNodeBs. A total of 60 cells configured in 20 eNodeBs makes up TA_2. Once a UE registers with the LTE network, the MME assigns a specific TAL, which consists of a group of TAs in the neighborhood to the UE's current position, as stated in the 3GPP technical specification [3]. As a result, when the UE travels out of its current TAL, it will be allocated a new TAL through the TAU procedure, which is initiated by the UE through its serving eNodeB [2]. In this thesis, the signaling overhead induced as a result of grouping cells to form TAs that can be used as a fundamental building block of TAL and the algorithms used for creating TAs are investigated by considering a real LTE network scenario.



TA	Number of sites	Number of cells
TA_1	6	18
TA_2	20	60

Figure 1. 1 TA and TAL concept in LTE

1.2. Problem Statement

The number of cells allocated to each TA is a compromise between the number of area updates and the volume of paging messages, or between uplink and downlink signaling burden, from the standpoint of planning and optimization [7]. Some of the factors that should be considered for TAU optimization are UE spatial density and mobility, whereas paging is dependent on traffic load characteristics of each cell [9]. According to Jessica and Mario [9], the signaling overhead arising from a static configuration of TAs can be reduced significantly by using Tracking Area List (TAL) for location management in LTE networks. For large networks, like Addis Ababa's LTE network the design of efficient TALs requires an underlying collection of TAs that have previously been planned. In the current configuration of Ethiotelecom's LTE network in Addis Ababa, TAL is not implemented. In this same LTE network, the MME stores the TA in which the UE is registered in the TA update (TAU) and uses it to page or track a UE in the network area, which means that the location management scheme in place, is the standard scheme. These TAs are designed and optimized for location management in the planning phase which implies that the TAs will no longer perform well as UE distribution and mobility patterns change over time. As a result, a TA redesign and reconfiguration may be necessary to reduce total signaling overhead. Nevertheless, such a method has not been investigated and applied for the Addis Ababa LTE network context. This thesis intends to carry on a TA design and optimization approach using distance-based clustering (K-means) and density-based clustering algorithms, Hierarchical Density-Based Spatial Clustering of Application with Noises (HDBSCAN), to produce optimal number of TAs that can used for efficient TAL implementation [10].

1.3. Objective

1.3.1. General Objective

The main goal of this thesis is to investigate the current TAs for signaling overhead and design new optimal number of TAs for Addis Ababa's LTE network so that the overall location management signaling load on the MME is kept to a minimum.

1.3.2. Specific Objectives

- Assess current Ethiotelecom's LTE network location management techniques;
- Study the current TAC configuration in Ethiotelecom's LTE network in Addis Ababa;
- Study the user's mobility pattern by transforming the call detail records (CDR) of each LTE network user into the drivable road network of Addis Ababa;
- Assess the signaling load of the current TAs based on the mobility data to produce baseline;
- Create TAs using distance and density-based clustering algorithms for the same LTE network topology and mobility data;
- Assess the performances of the new TAs obtained as a result of the distance-based clustering algorithm against the baseline;
- Assess the performances of the new TAs obtained as a result of the density-based clustering algorithm against the baseline;
- Identify TAs configuration, that minimizes the overall signaling load and make recommendations.

1.4. Literature Review

Many technologies have been discussed about location management and many researchers have investigated the induced signaling overhead as a result of users' mobility. TAU and paging, in particular, have been identified as two key components of the location management method. Several investigations, like in [11], [12], and [13] potential alternative strategies for TAU signaling overhead minimization derived from traditional techniques such as timing, speed, and mobility for triggering the TAU procedure, thereby reducing signaling overhead. The TAL concept has intrigued scholars from both academia and the industry due to its preferred adaptability. TAL configuration study can be divided into two main categories: static and dynamic configuration. Several researchers looked at the TAL in various methods, with some using traditional optimization approaches connected to the notion of TAL in previous works [5], [13], [14].

In Razavi et al. [14], the researchers planned and evaluated by comparing a new TAL design approach to the conventional TA concept. The researchers also confirmed the significant dissimilarities in cell distribution between TA and TAL in a static mode. The same researchers suggested a dynamic TAL design in associated work [13], in which the cells are dynamically allocated to the system's existing TALs configurations. The rule of thumb approach was developed to adapt to the design of TAL in order to minimize the signaling overhead caused by TAU and paging, and it was used to allocate cells to tracking area lists. The study's findings show that the proposed dynamic approaches outperform the standard TA and static TAL in terms of performance results. Furthermore, the authors proposed a study that would use a linear programming model to minimize signaling load [5]. The TAU and paging signaling overhead regarding TAL creation were explored in two studies by Aqeeli et al. [8], [11], to link the MME with TAL. In both experiments, a centralized and a distributed MME technique were proposed. Each TAL construction is linked to one MME in the centralized system, but each TA that characterizes each cell is linked to one MME in the distributed scheme. Both types of research examine the implications of having various MME distributions throughout the system and the overall signaling overhead consequence. The first work [11], in contrast, provided a static technique for allocating cells to tracking area lists in a fixed period. When compared to the distributed MME technique, the centralized MME technique provides a more streamlined signaling overhead. The second study [8] presented an intelligent dynamic procedure that restructures cells among the TALs in a preset time based on a smart selection scheme. The smart selection strategy aids in the inclusion of cells in the same tracking area list that are more likely to have big connections between them. The research discovered a substantial improvement in total signaling overhead.

Jessica and Mario [9], discuss that the design of TALs requires an underlying set of TAs that has already been planned, though some TAL optimization algorithms assume each TA is just made up of one cell. However, in networks using current standards, assuming a TA to be only one cell may not be ideal. As a result, a K-means clustering algorithm is utilized to generate TAs to find the best Voronoi distribution of TAs that will be employed by the TAL optimization [9].

1.5. Methodology

The essential contextual information about TA design and optimization in general and distance-based clustering as well as density-based clustering algorithms are covered through literature reviews from journals and books.

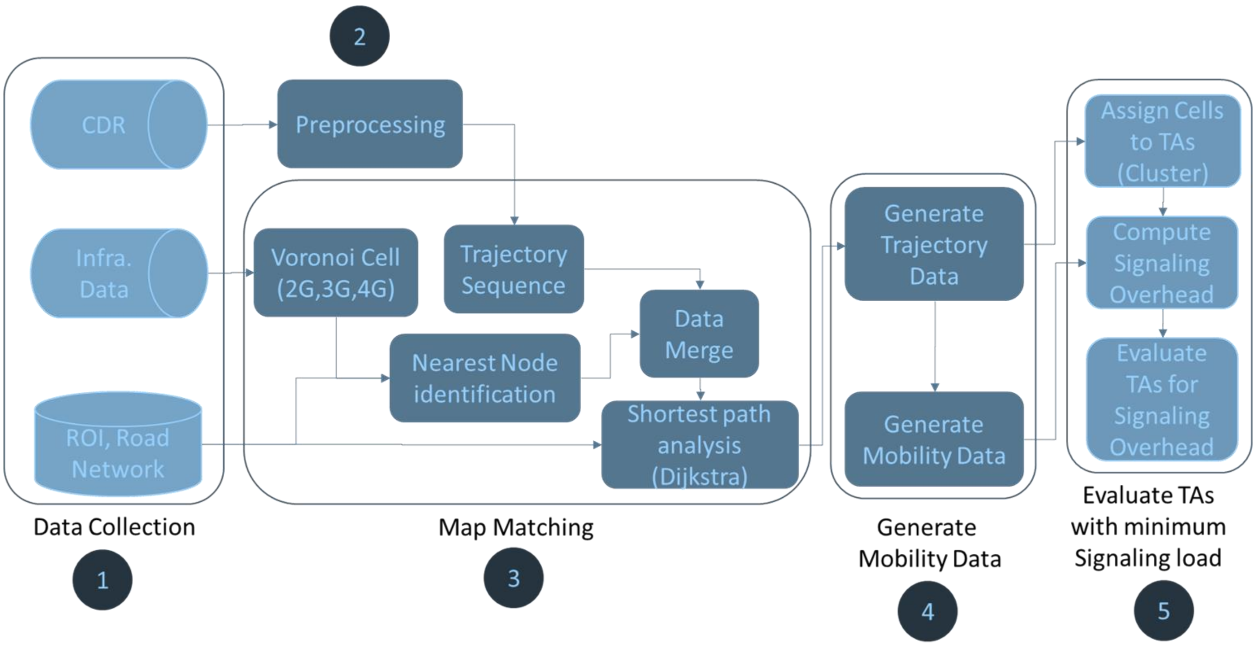


Figure 1. 2 Analysis Procedure (Methodology)

Call detail records data is collected from Ethiotelecom Customer Relation Management (CRM), infrastructure data (LTE network design data) was collected from the wireless Engineering division, and for map matching, Openstreet map data is downloaded and used for filling the location gaps observed in the CDR data. A laptop computer with Python environment with libraries: Geopandas, OSMnx, NetworkX, Shapely, scikit-mobility and other necessary tools are used. The methodologies followed to accomplish the general objective as well as the specific objectives of this thesis are depicted in Figure 1.2 below.

1.5.1. Data Collection

To acquire the required data for this thesis, an official supporting letter from the Addis Ababa Institute of Technology (AAiT) is submitted to the responsible personnel in Ethiotelecom.

Accordingly, LTE users' voice calls CDR data for twenty days from Ethiotelecom's CRM system were collected in XML format. Similarly, the cell plan, for 2G, 3G and LTE networks, was collected from the respective division of Ethiotelecom. Addis Ababa's road (drivable) network data was downloaded from the Openstreet map.

1.5.2. Data Preprocessing

The required features from the CDR data are shown in Table 1.1 and any record with missing values of any of the parameters is dropped. By the same token, the required parameters of the cell plan data are shown in Table 1.2. To get the area of interest, Addis Ababa's administration boundary shape with geometry parameter is required, as shown in Table 1.3. The road network data parameters for road intersections (nodes) and road segments (edges) are shown in Table 1.4 and Table 1.5 respectively.

Table 1: 1 Call detail records (CDR) parameters

Feature	Description
Caller id	Call originating user id
location id	Call originating user location id
Start time	Originating call start time (timestamp)

Table 1: 2 Infrastructure parameters

Feature	Description
latitude	Site latitude
longitude	Site longitude
location id	Cell identity
Azimuth	Angle between cell antenas
site id	Site identity

Table 1: 3 Addis Ababa's city admin boundary shape

Feature	Description
geometry	Geometry of Addis Ababa's city admin boundary shape

Table 1: 4 Openstreet map data (road network data – nodes)

Feature	Description
osmid	A road network intersection (Node) identifier
latitude	Site latitude
longitude	Site longitude
geometry	Geometry of the road intersection

Table 1: 5 Openstreet map data (road network data – segments)

Feature	Description
u	Source node osmid
v	Destination node osmid
length	Length between source node and destination node
geometry	Geometry of the road segement

1.5.3. Shortest Path Analysis (Map Matching)

The CDR data is recorded when a user initiates a call and the cell with which it was attached at the beginning of the call is recorded as location id. The location id in the CDR doesn't show the whereabouts of the user in the idle mode. To make an educated guess about the user's idle mode mobility path in the network a map-matching technique is employed, which included the following sub-steps:

- Creating the Voronoi tessellation of the cellular network
- Creating Trajectory sequences from the CDR
- Searching the nearest road network intersection (node) to the centroid of the Voronoi cells
- Aligning the trajectory sequence created for the CDR data with the location information in the road network data
- By means of the bidirectional Dijkstra algorithm [15], the shortest path that would have been followed by the UE is calculated and the location information of the road network intersection (nodes) is populated to form the dataset for the clustering and flow computation steps.

1.5.4. Generating Mobility Data

The UE traces or trajectory data is generated using the map-matched CDR data and the outflow (number of users moving out of a cell to the other cells) is computed based on the transition probability computed from the trajectory data and the Voronoi cell tessellation of the LTE network in Addis Ababa.

1.5.5. Signaling Load Computation and Evaluation

To create TAs (group of cells), a distance-based clustering algorithm (Kmeans) and a density-based algorithm (HDBSCAN) are considered. The cells membership information for each new TA and the outflow information are repeatedly used to compute the signaling load, till TAs that produces the minimum overall signaling load are identified. This procedure is recognized to be a repeated heuristic algorithm. The minimum overall signaling load computed for TAs created with distance-based clustering algorithm (Kmeans) as well as the minimum signaling load computed for TAs created with the density-based clustering algorithm (HDBSCAN) in each repeated search is compared against the baseline signaling load of the current TA configuration of Addis Ababa's LTE network. The performance evaluation is measured as a percentage improvement to the baseline signaling load.

1.6. Scope and Limitations

This research is focused on Addis Ababa city's LTE network based on sample CDR data taken from the CDR of randomly sampled 9,307 LTE users over a 20-day long observation. Allowing for a larger number of users and user behavior profiling may yield a different result.

1.7. Contributions

The important contributions of the thesis are briefly put as follows.

- Shows how to use a heuristic approach for solving the minimization of the overall signaling overhead in the LTE network of Addis Ababa, close to optimality, by using two separate clustering techniques embedded within a repeated heuristic algorithm.

- Exploring the challenges in cell assignment to TA based on distance and density-based clustering algorithms
- A performance comparison of two suggested approaches for assigning and allocating cells to TAs for the LTE network in Addis.
- The results, TAs with overall minimal signaling load, can be used for TAL assignment based on the predicted mobility pattern of each UE.

1.8. Thesis Outline

The remaining portion of the thesis is structured as follows. In Chapter two, some of the relevant previous research on location management schemes are explained. Second, the standard TA scheme is explained and at the last section, the total signaling overhead formulation employed all over this thesis work is presented. Chapter three presents the nearest node search algorithm employed to match the centroids of the Voronoi cells to the road network nodes, shortest path analysis algorithms to produce the UE’s traces, UE’s distribution and the UE’s mobility data, specifically the outflow. It also explains mobility analysis, used to produce the outflow data. Chapter four presents the LTE network location management signaling load analysis approaches in detail. In Chapter five, results and discussion are presented. Chapter six, the conclusion and recommendations of this thesis work are presented.

Chapter 2

Tracking Area in LTE Networks

This chapter provides the context and foundational materials for optimizing location management signaling overhead in LTE networks. Tracking area update signaling and paging signaling are two types of location management signaling which are dependent on the design and configuration of tracking areas. The TA design and optimization principles considered throughout the thesis, is presented.

2.1. LTE Architecture

Understanding the LTE system architecture and its implementations is critical for realizing location management functionalities. The LTE core network architecture in charge of providing entire mobile-broadband services to the underlying Evolved Universal Terrestrial Radio Access Network (E-UTRAN) is known as the Evolved Packet Core (EPC) [15].

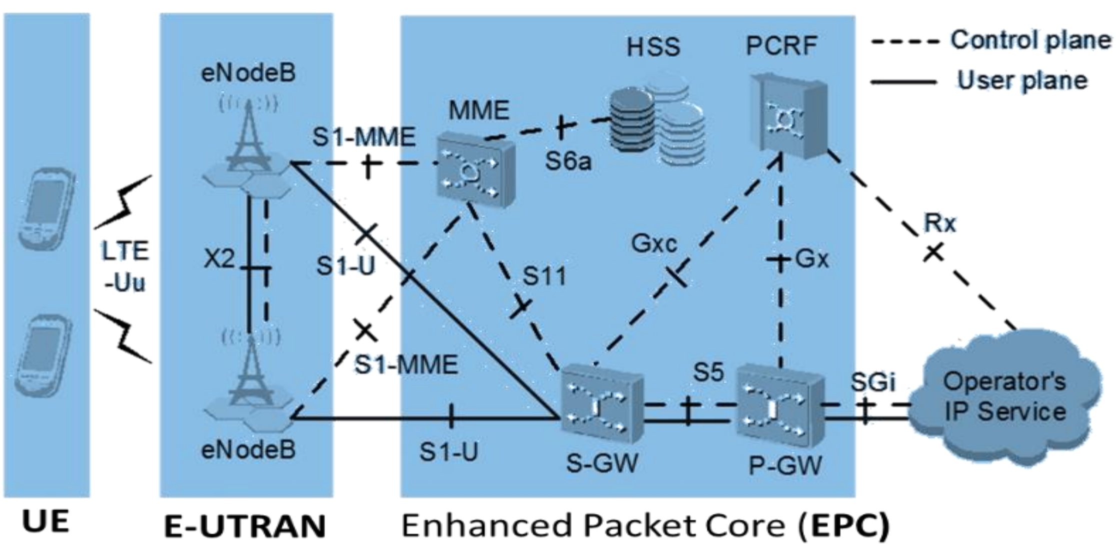


Figure 2. 1 Architecture LTE networks (EPC and EUTRAN)

According to 3GPP specifications [3], the EPC is made up of several components, including the Serving Gateway (S-GW), Packet Data Network Gateway (P-GW), Policy and Charging Rules Function (PCRF), Home Subscriber Service (HSS), and MME. The MME, also known as the EPC's control plane, is in charge of mobility management, among other functions.

Authentication, handling of idle to active transitions, handover, TAL management, and paging are just a few of the control-plane functions supported by the MME [3]. The TAU and Paging techniques (managed and controlled by the MME) and the associated signaling overhead are the focus of this thesis. These two location management procedures are critical because they must track and locate each user equipment (UE) as it moves within the network coverage area.

2.2. Basic Technical Terms

This section explains the technical definitions created by 3GPP in Release 9 [3]. The terms listed underneath are used throughout the thesis and are covered here to provide context for the complete research.

- An end-user device **(UE)** is any device used directly by an end-user to communicate over an LTE network.
- A **cell** is a radio coverage area identified by the identification of a base station (eNodeB, in the case of LTE networks).
- The **Tracking Area (TA)** is defined as a region in which a user can freely move without updating the MME. The term TA is used in LTE networks. The network provides a list of one or more TAs to the user. In certain operating modes, the UE may freely navigate through all of the TAs in the list without updating the MME.

2.3. Location Management

There is a huge amount of information and material accessible, on cellular network location management [2], [6]. All issues and challenges during location area (LA) and routing area (RA) planning and optimization can be applied to TA research. Some location management solutions have been proposed in the literature. The vast majority of these strategies have been studied and classified in [6], [16]. This section aims to compile a list of the schemes that have been thoroughly researched. There are two types of update systems: location area update schemes and paging schemes.

2.3.1. Tracking Area Update Schemes

TAU begins with a user-generated update message sent over the uplink control channel, followed by signaling that updates the database. Because of the use of network bandwidth and core network communication to update location databases, each TAU is an expensive operation. Based on the 3GPP LTE specifications [3], Alsaeedy and Chong [2] summarized and introduced Table 2.1 to show the signaling overhead of TAU and paging corresponding to each involved network entity (network entities are shown in Figure 2.1), measured in the number of required messages M for each network element. For example, if a mobile UE needs to initiate this procedure by reporting a location change, the overall cost per TAU will be 40 M. When a network, specifically the MME, needs to initiate this process for delivering UE-specific data packets, the overall cost per Paging will be 33 M.

Table 2. 1 Signaling cost of TAU and Paging in number of messages (M) [2]

Network Element	TAU (M)	Paging (M)
UE	9	5
eNodeB	14	8
MME	7	8
S-GW	4	6
P-GW	4	4
PCRF	2	2

There are numerous methods for reducing the number of update messages received from users, which can be divided into two categories: static and dynamic TAU implementations. In static schemes, tracking area updates are triggered based on the LTE network architecture, but in dynamic schemes, tracking area updates are activated depending on the user's call and mobility patterns. Because static techniques are independent of user attributes, they allow for efficient implementation and low computational requirements. Unlike static schemes, dynamic schemes typically necessitate online data collecting and processing, which consumes a large number of CPU resources. Dynamic systems, on the other hand, have a higher amount of signaling overhead reduction than static methods. As a result, careful design is required for dynamic schemes for the network to successfully support the computation [2], [6]. Figure 2.2 depicts the summary of the location management schemes reviewed for this work.

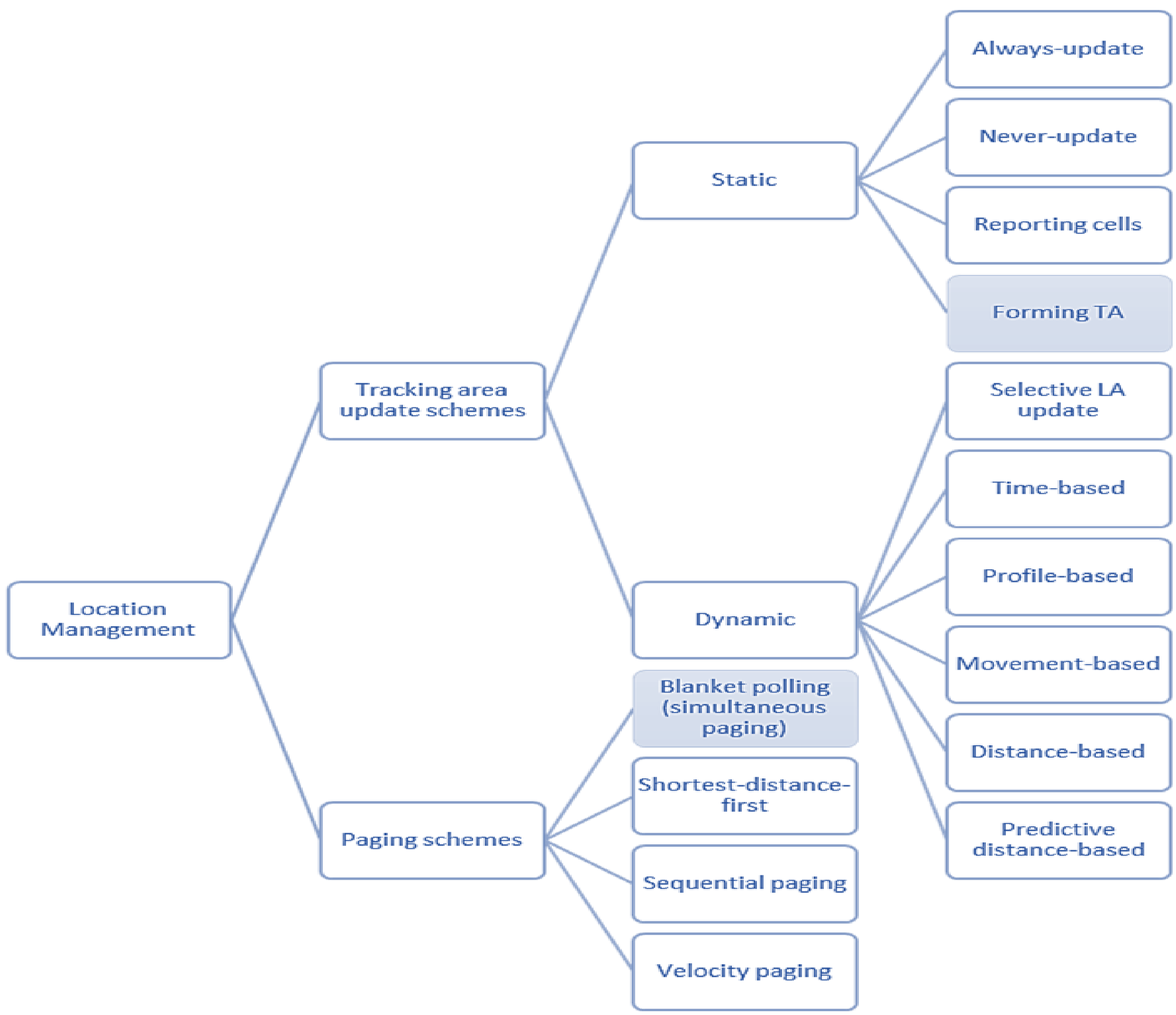


Figure 2. 2 Location Management Schemes

A brief explanation for TAU schemes and paging schemes is presented in Table 2.2 and Table 2.3, below.

Table 2. 2 Tracking Area Update Schemes

Type	Update Scheme	Explanation
Static	Always-update	As soon as a user enters a new cell in this scheme, it updates its location to the MME. No paging is required because the network has complete knowledge of the user's location. For consumers with low mobility and high call arrival rates, this strategy works well. Due to the high number of TAUs, this scheme is rarely employed.
	Never-update	The user never changes its location information (i.e. same TA) in this scheme, hence there is no location update overhead. However, for large-scale networks and networks with high call arrival rates, it results in excessive paging. This scheme is hardly seldom utilized as well.
	Reporting cells	The user's location is only updated in the MME when a user visits one of the designated reporting cells. A search must be undertaken in the region of the last reporting cell from which the user has updated its location to paging a user [6]. It is impossible to assign an optimal reporting cell arrangement without taking user movement into account.
	Forming TA	When the user changes a TA in this method, the user's location is updated. A user's paging will take place within the TA in which the user is positioned. This scheme is known as the standard update scheme, and it is the thesis's presumed scheme.
Dynamic	Selective LA update	The TAU is not conducted every time the user crosses a LA border in this system. Because the user will only be in such LAs for a short time, the TAU process at those LAs can be avoided [7], [11].
	Time-based	The user updates its location at fixed intervals to the MME, in this scheme. The time interval can be improved per user to reduce the number of update messages [2].

	Profile-based	In this method, the network keeps track of each user's profile and each user's profile is updated on a regular basis [7]. The profile contains a list of the most likely TAs in which the user is located at various times and the TAs on the list are sorted by the likelihood of finding a user, from most likely to least likely.
	Movement-based	The user's location in this scheme is updated after a certain number of TA boundary crossings with other cells in the network. Based on the particular movement and call arrival behavior of each user, the boundary-crossing threshold can be optimized [2].
	Distance-based	In this scheme, the user updates its location when it moves a certain distance away from the cell where it last updated its location. Based on the user's specific movement and call arrival pattern, the distance threshold can be optimized per user [2].
	Predictive distance-based	Based on location and speed data, the network calculates the probability density function of the user's location in this manner. When the user's distance from the projected location exceeds the threshold, TAU is performed [2].

2.3.2. Paging Schemes

The network determines a specific user's position at the cell level through paging. A polling cycle refers to each phase in the process of identifying a user's location in the network [3]. Polling signals are broadcast via the downlink control channel to all cells in a TA, where a user is anticipated to be present during each polling cycle. Only the called user sends a response message back over the uplink control channel after hearing the paging message. Radio bandwidth is consumed during the paging operation and the paging overhead is proportional to both the number of polling cycles and the number of cells polled every cycle. There is a timeout period in each polling cycle, and if the user is not located within that time frame, a new group of cells will be chosen in the following polling cycle and the maximum paging latency is determined by the number of polling cycles that can be used to locate the user. Table 2.3 shows the summary of the paging schemes.

Table 2. 3 Paging Schemes

Paging Scheme	Explanation
Blanket polling or Conventional paging	All cells in the user's TA are paged at the same time under this technique which does not require any additional information about the user's position, and it is the most feasible and widely utilized scheme in today's networks. conventional paging scheme.
Shortest-distance-first	In this technique, the network pages the user by starting at the last cell where the user changed its location and working outward in shortest-distance-first order.
Sequential paging	The user is paged successively in sub-groups of cells in the LA in this technique, where the sub-groups are arranged in order of their assessed chances of containing the user.
Velocity paging	Users are categorized in this scheme depending on their velocities at the time of location updates where paging area is generated dynamically depending on the user's most recent TAU time and velocity class index [2].

Although the selective TAU and paging systems presented in this and the previous section reduce signaling overhead, their use necessitates system modifications and the collecting of additional user data. As a result, the standard scheme is still commonly used.

2.4. Location Management in Addis Ababa's LTE network

The location management system for Addis Ababa's LTE network has three MME (i.e. MME_1, MME_3, and MME_3) which are connected as mesh and used as an MME pool as shown in Figure 2.3.

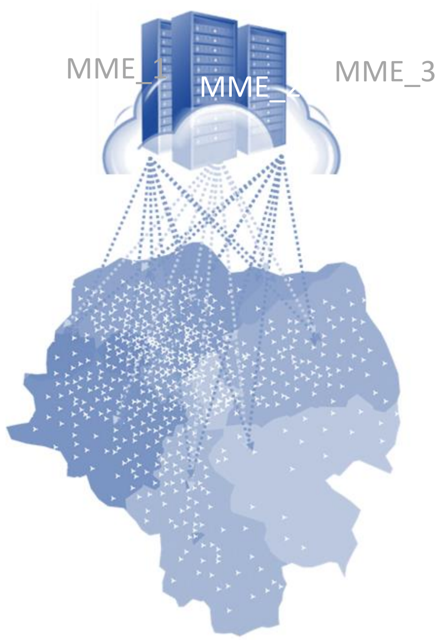


Table 2. 4 Number of cells per TA

TA	Site ID count	Cell ID count
TA_1	128	384
TA_2	123	378
TA_3	56	174
TA_4	111	382
TA_5	46	138
TA_6	101	303
TA_7	72	246
TA_8	96	288
8	**650**	**2206**

Figure 2. 3 Core (MME) and radio network topology of Addis Ababa's LTE network

In the MME Pool networking mode, one eNodeB can be connected to multiple MMEs having the following advantages:

- Multiple MMEs in the MME Pool share the load of the network which improves the utilization of resources of the core network and saves investment in equipment.
- Data can be backed up among the MMEs in the MME Pool, as a result, it improves the reliability of the network.
- The number of inter-MME location updates in the MME Pool and signaling traffic are reduced, as the MME has the same location information.

Ethiotelecom adopts a conventional paging [15] scheme for paging and area-based strategies for a TAU [15] and to search the called UE the MME broadcasts polling messages to a TA using the last updated information registered. Addis Ababa administration boundary's LTE network has 8 TAs, 650 eNodeBs, and 2206 cells. The number of eNodeBs and cells per TA is shown in Table 2.2, and each TA serves hundreds of cells.

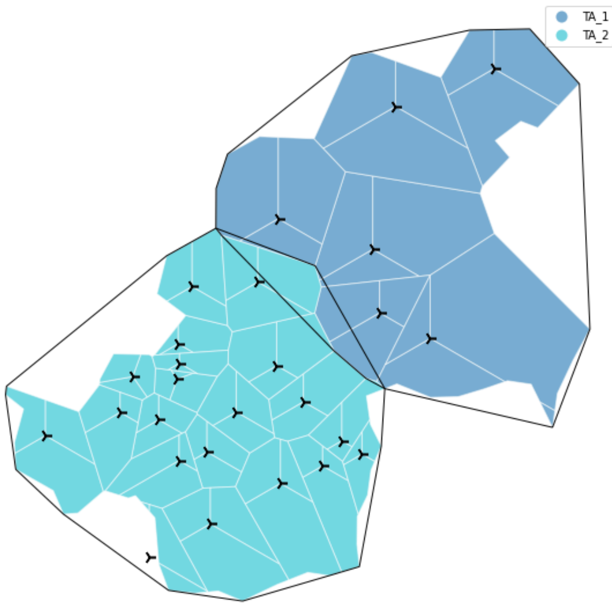
Table 2. 4 Number of cells per TA

TA	Site ID count	Cell ID count
TA_1	128	384
TA_2	123	378
TA_3	56	174
TA_4	111	382
TA_5	46	138
TA_6	101	303
TA_7	72	246
TA_8	96	288

It is also important to note that for each eNodeBs (including cells configured in it) a single TA is configured i.e. each eNodeBs site in the administration boundary of the city of Addis Ababa, is taken as a cell for planning the TAs and the distribution of the cells to TAs [15].

2.5. TA Design Optimization

The fundamental design task in the typical architecture of TAU and paging is the generation of TAs, to minimize total signaling overhead. Paging is almost eliminated when TAs are very small (e.g., one cell per TA), but this induces excessive TAU, whereas TAs that are excessively large have the reverse effect. As a result, tracking area planning and optimization task's sound goal is to achieve the best possible balance of TAU and paging signaling messages volume. Tracking area planning has a basic trade-off as depicted in Figure 2.4. For example, TA_1 has 6 eNodeBs in which 18 cells are allocated and TA_2 has 20 eNodeBs in which 60 cells are allocated. Assuming a UE has the same incoming call connection rate, the paging overhead as a result of this UE's presence in TA_2 will be much higher than the paging overhead as a result of this UE's presence in that of TA_1. The TAU is not required as the UE roams in the 60 cells in TA_2, however TAU is not required only when the UE the 18 cells in TA_1.



- Small TA: More TAU, less paging
- Large TA: Less TAU, more paging

TA	Number of sites	Number of cells
TA_1	6	18
TA_2	20	60

Figure 2. 4 An illustration of the TAU and paging trade-off in LTE networks.

Sara M. [4] tackled the trade-off problem via mathematical programming. They demonstrated an integer programming model and a cutting plane algorithm, as well as the optimality of a 38-cell GSM network. Because the trade-off problem is NP-hard, heuristic techniques such as insertion and exchange heuristic [4], simulated annealing [16], and evolutionary algorithms [7] are commonly used to solve huge networks. Alsaeedy and Chong [2] compares and contrasts a variety of heuristic algorithms for LA design in terms of optimality and computational effort. The researchers of [17] discuss cell-to-switch assignment for load balancing in addition to LA design. The researchers of [8] present an integer programming model for this problem as well as a graph-partitioning heuristic-based solution strategy. The researchers of [8] predicts TAU and paging requests using simulation techniques.

For location management signaling overhead investigation and TA design, this thesis uses the conventional TAU and paging approach as this is the scenario for the case study LTE network. This means that when a UE crosses the TA boundary, a TAU message is sent and paging is initiated in all cells of the TA to which the UE is currently registered.

2.6. User Equipment States in Mobility Management

In LTE, User Equipment (UE) refers to any device used directly by an end-user to communicate over the LTE network, which could be a hand-held phone, a laptop computer, or any other device with a mobile broadband adaptor can be used as a UE. The UE can be in one of these three states, as shown in Table 2.5, in terms of mobility [3].

Table 2. 5 User Equipment States in Mobility Management [3]

State	Explanation
LTE-Active	The network recognizes the UE's cell, and the UE can send and receive data from the network. Active UEs do not require TAU or paging.
LTE-Idle	The UE is in power-saving mode in idle mode, does not notify the network of each cell change, however, the network can pinpoint the UE's location down to a few cells (forming a TA).
LTE-Detached	In this mode, the UE is either switched off or in a transitory condition, in which it is searching for and registering with the network.

The UE is often in the LTE-Idle state, and the MME knows which TA the UE was last registered in. Cell load and cell handovers are usually the only realistic data available from a cellular network which actually represents the active UEs. If idle UEs have the same mobility behavior as active UEs, cell load and handover statistics can be a reliable estimate of their location and movement. Network simulation and analyzing traffic density on road networks across neighboring cells are two methods for evaluating the mobility behavior of idle UEs [18]. As the terms cell load and handover are commonly used to refer to active UEs, it is not used in this thesis. However, the CDR of each UE and the drivable road network graph data Addis Ababa is used to synthesize the distribution and mobility pattern of idle UEs.

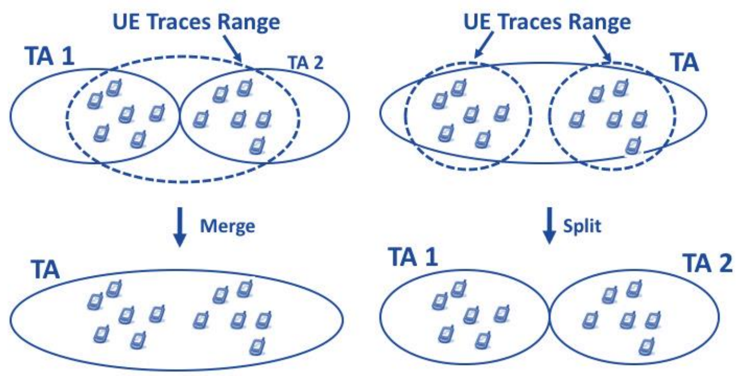


Figure 2. 5 Merge and split of TAs. [4]

A UE trace is described as a UE's cell-to-cell mobility pattern and call arrival pattern over some time, and it helps toward minimizing signaling overhead and optimize TA design [7], [4]. Considering the range of UE traces movement for the selected area as shown in Figure 2.5, the UE-traces range in the left figure shows that TA1 and TA2 should merge. The separation of UE traces in the right image illustrates that the signaling overhead is decreased by breaking the TA into two smaller TA. From this example, it can be deduced that even a preliminary estimate of the UE traces might be beneficial in planning and optimizing TAs.

Chapter 3

Mobility Analysis for Addis Ababa's LTE Network

In this chapter, the methodology employed including the data collection, features selection, and data preprocessing are presented. The chapter discusses how the trajectory gap that might be present in the CDR data of some individual UEs, is reconstructed through the shortest path analysis on the drivable road network within the administration boundary of the city of Addis Ababa. It also presents the mobility analysis and the summary of the resulting data used as an input for TA design and optimization analysis in Chapter 5.

3.1. Data Preprocessing

3.1.1. Sampling Call Detail Records (CDR)

The features in the LTE users CDR (voice) data are many, however only the required parameters are shown in Table 1.1, are requested and collected for the study period of 20 days. For security reasons, the user's identity in the CDR data is anonymized before data handover to the researcher. The format and the size of the CDR data are extra markup language (XML) and 38 GB respectively. So, file splitting and parsing tools in the python environment, like *fsplit* and *xml2csv* are used to get manageable data in the appropriate format for the tools (*pandas*) and the machine (Laptop with Windows 10 & 8 GB RAM) used.

The voice service of the LTE network users uses a fallback mechanism [15], hence the Voronoi tessellation of the cell plan of the 2G network, the 3G network in the city is constructed (section 3.1.2) to get the location information of the UEs. As the LTE network covers areas outside of Addis Ababa's administration boundary, the city's geometry shape file (section 3.1.2) is used as a clipping geometry to extract the cells within the city and drop the rest. These cells are used to extract the CDR of the LTE users from which a total of 930,774 unique users are identified. Once more, due to the processing capacity of the laptop computer used for this research and the huge CDR data size, 1 % of the total unique users is randomly sampled, resulting in 9,307 users. The

CDR of each user in the sample is extracted from the total CDR data which returned 574,610 records. This extracted CDR data is used for the shortest path analysis, explained in section 3.2.

3.1.2. Cell Plan (Voronoi Tessellation)

The voice service for the LTE network under study uses fallback mechanisms, which implies the location information in the CDR data is either the 2G network or the 3G network in the same city. The location id in the CDR data and the location identity in the cell plan data are used for adding the latitude and longitude parameters to the CDR data. To approximate the coverage area of each of the cells of the network (2G, 3G, and LTE), the Voronoi tessellation is constructed, using the features in the cell plan, shown in Table 1.2 and the Geopandas package [19]. Figure 3.1, shows the LTE Voronoi cells, as an example.

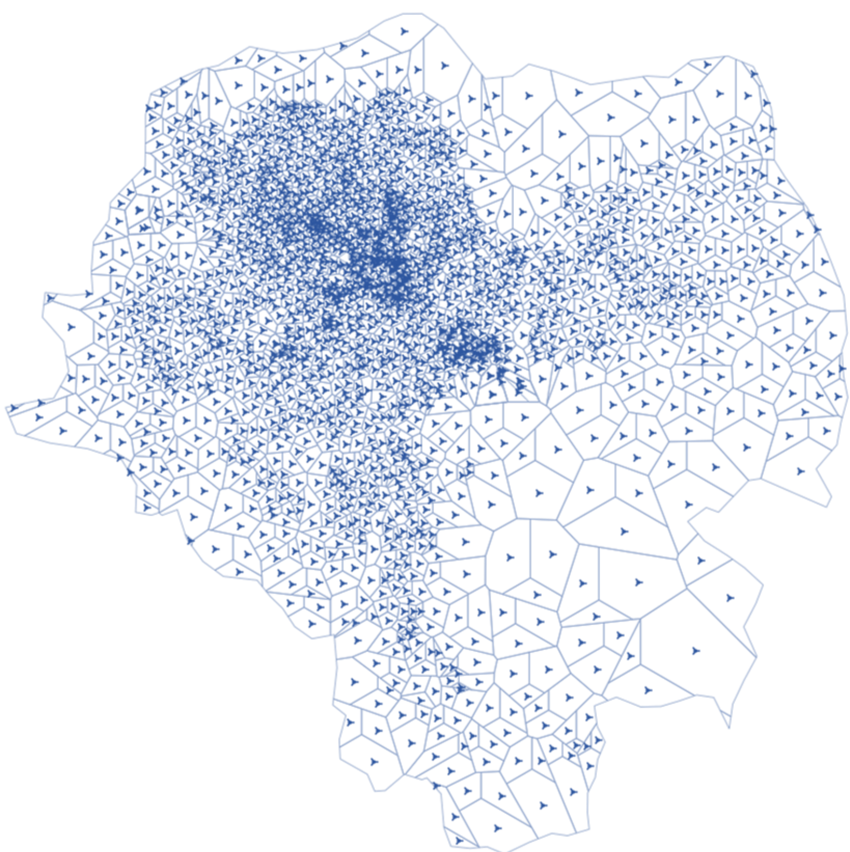


Figure 3. 1 Voronoi Tessellation for LTE network in Addis Ababa's Administration boundary

3.1.3. Addis Ababa's Admin Boundary Shape

The geometry of the administration boundary of Addis Ababa is downloaded from [20]. The important geometry parameter, shown in Table 1.3, is used for clipping the Voronoi cells geometry (to approximate coverage area) of the cells and identifying the cells within the city of Addis Ababa. This same geometry is used for clipping and downloading the drivable road network of the city of Addis Ababa from the Openstreet map [21].

3.2. Shortest Path Analysis (Map-Matching)

3.2.1. Road Network Data

The drivable road network data within the administration boundary Addis Ababa is downloaded from the Open street map [21] using packages of OSMnx [22]. This data has a lot of features, though the required parameters for shortest path analysis are extracted. The road intersections (nodes) parameters are shown in Table 1.4, and that of the road segments are shown in Table 1.5.

3.2.2. Nearest Node Analysis

Using the road network graph data and the centroid of the Voronoi cell, the nearest nodes (i.e. road network intersection) to each of the cells centroid is identified among the ten nodes in the neighborhood of the cell centroid [22].

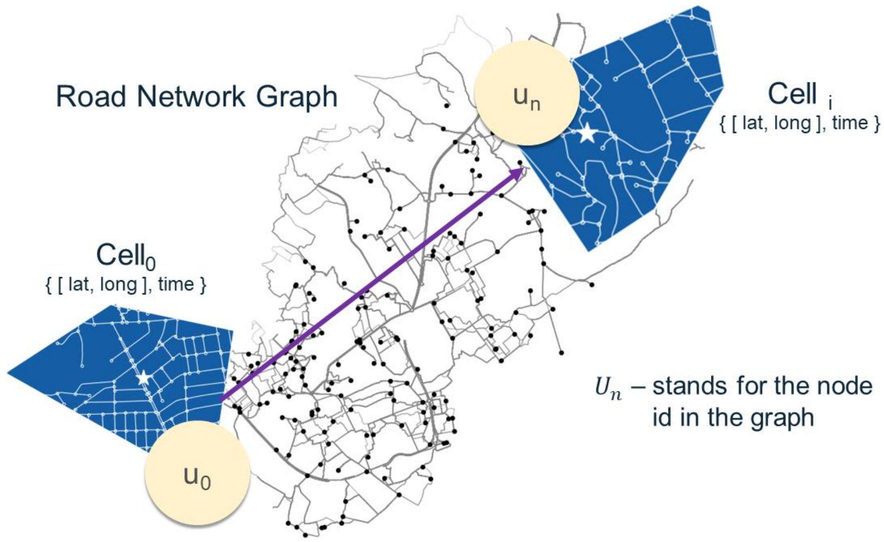


Figure 3. 2 Voronoi Cell centroid mapping to the road network intersections.

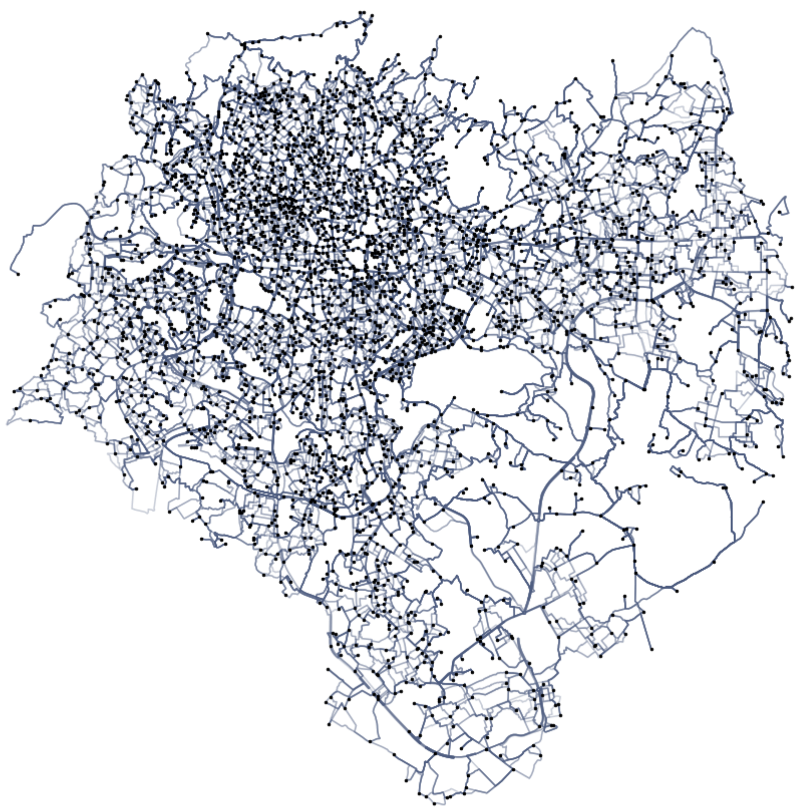


Figure 3. 3 Road Network Graph in the city for Shortest Path Analysis

The final graph, representing the Voronoi cell centroids of the 2G and 3G network is reconstructed as shown in Figure 3.3. The white dots represent the road intersections (nodes) that a UE starts or destines a trajectory. In the same graph, the blue lines symbolize the road network segments (edge) that a UEs might move on from a source node to the destination node. This graph is used for the shortest path analysis to reconstruct the UE traces from the corresponding traces in the CDR data.

3.2.3. Rebuild UE's Trace

UE traces of sampled CDR data are known to show both spatial and temporal sparsity for some LTE users. To fill this gap, each UE trace observed in the sampled CDR data is substituted by the information on waypoints (nodes in the road network) where the UE may have passed through Addis Ababa's road network diagram. Estimates are obtained and calculated using the shortest path analysis algorithm known as Bidirectional Dijkstra [23]. The Euclidean distance between two consecutive points calculated and the path (from source to destination) with the least cost (highest probability) is selected among other alternative routes in the road network graph. The location information of each node on the shortest path is entered for each UE trace in the CDR. Assuming the average UE velocity between the start and end location of each trace, the arrival time to each road network intersection along the shortest route is calculated, and the location information and time of the trace observed in the original CDR data is used to replace information. Some records of the reconstructed UE traces of all UEs in the sample CDR, is shown in Table 3.1 below.

Table 3. 1 UE traces for the LTE users in Addis Ababa (Reconstructed sample)

	Caller id	**Start time**	**u**	**Location id**	**Longitude**	**Latitude**
5482540	fffbb3c5c2e8	5/1/2021 19:25	262627495	1133	38.76247568	8.988843786
5482541	fffbb3c5c2e8	5/1/2021 19:37	6307707827	1536	38.75012871	8.992404156
5482542	fffbb3c5c2e8	5/1/2021 19:41	299009340	112201	38.75883528	8.991407367
5482543	fffbb3c5c2e8	5/1/2021 19:43	560335266	1299	38.75072041	8.998068166
5482544	fffbb3c5c2e8	5/1/2021 19:49	48613467	1356	38.74941968	9.008796636

3.3. UE Distribution

The total number of UEs at cell i (denoted by N_{UE_i}) is defined as the sum of the individual UE's proportion of time a user spends in cell i, Equation (3.1). Waiting-time (or inter-time), which is defined as the time between two consecutive points in a UE's movement, is computed in a scikit-mobility library in Python [24], [25] to get the time an individual UE spends in cell i. The calculated number of UEs in each of the cells is tabulated in Appendix B.

$$N_{UE_i} = \sum_{\forall users}\left(\frac{Total\ time\ at\ Cell_{\ i}\ of\ a\ UE}{\sum_{j\in N_{cell}} Total\ time\ at\ Cell_{\ j}\ of\ a\ UE}\right) \quad (3.1)$$

Because UEs are more likely to exist in cells with a higher number of UEs [4], [26]. Equation (3.2) calculates the probability ratio of cell i being the location where a UE can occur during the specified time period.

$$P_{loc}(i) = \frac{N_{UE_i}}{\sum_{j\in N_{cell}} N_{UE_j}} \quad (3.2)$$

Equation (3.3) describes the up-scaled distribution of UEs in the LTE network as a function of the total number of unique LTE users identified in the CDR data.

$$\boldsymbol{N_{UE_i}} = (930774 * P_{loc}(i)) \quad (3.3)$$

Figure 3.4 depicts the distribution of UEs in Addis Ababa's LTE network. The sites are represented by Voronoi tessellation of each, and the numbers at the centroid of each Voronoi cell are set in proportion to the total number of UEs in the reconstructed UE's traces scaled by the total number of unique LTE users identified in the entire dataset of the 20 days' duration.

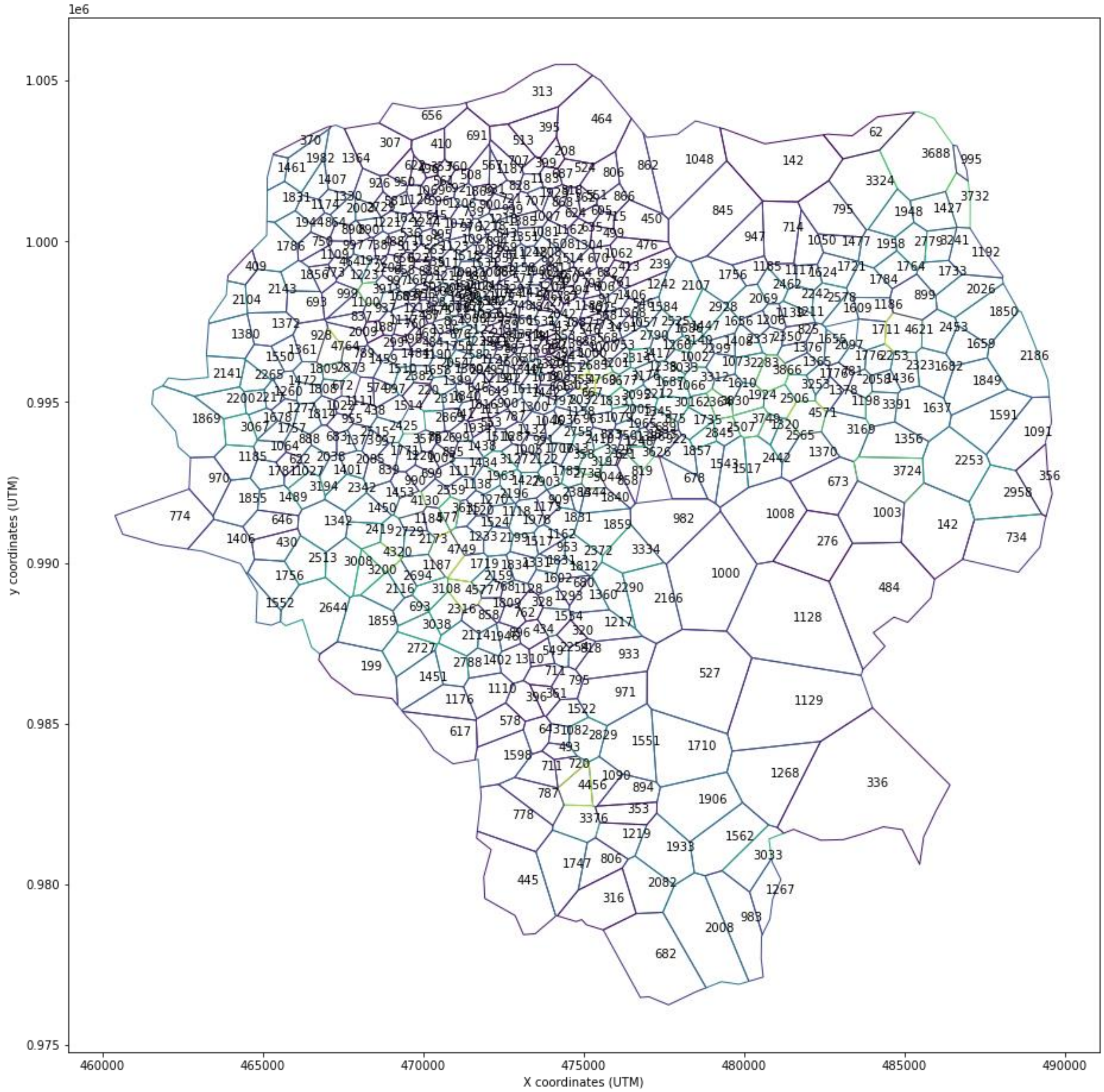


Figure 3. 4 UE distribution for the LTE network in the city (eNodeB level)

3.4. Number of Crossings (Flow)

The probability of finding a UE at cell i calculated in section 3.3. above can't be used for calculating the crossings, as it doesn't give us how many of the UEs estimated to move out of cell i moved out to cell j. So, the principle, users (UEs) have a preference to move to neighbors with high flow or crossing values [16], [26] is employed. Equation (3.4) calculates the likelihood of a

UE moving from cell i to cell j, assuming that cell i ≠cell j. If cell i = cell j, Equation (3.4) calculates the likelihood of a UE staying in cell i. The transition probability from cell i to cell j can be defined as:

$$\boldsymbol{H_{ij}} = \frac{Total\ flow\ from\ cell\ i\ to\ cell\ j}{Sum\ of\ all\ flows\ from\ cell\ i\ to\ cell\ j\ for\ all\ N_i} = \frac{h_{ij}}{\sum_{j \in N_i} h_{ij}} \qquad (3.4)$$

where N_i denotes the set of cells having a non-zero flow values from cell i.

Table 3. 2 Transition Probability

destination origin	111001	111002	111003	....	117140	117141	117144
111001	0.146906637	0	0	....	0	0	0
111002	0	0.093620648	0.00109569	....	0.000121743	0.000852204	0
111003	0	0.005999077	0.053991694	....	0.000922935	0	0
....	....	....	....	....	....	....	....
117140	0	0	0.000062243	....	0.074442923	0.01188846	0
117141	0	0	0	....	0.013745704	0.347186426	0
117144	0	0	0.000290444	....	0	0	0.211733953

Here, the Voronoi tessellation (eNodeB level) of the LTE network is used for aggregating the number of crossings (flow), from cell i to cell j from the reconstructed UE's traces, using scikit-mobility. Table 3.2, shows the transition probability, for instance, the likelihood that a UE at cell '111002' to stay at the same cell is 0.093620648, while the transition probability to cell '111003' is 0.00109569. The complete transition probability matrix data is included in the CD submitted together with this thesis document, as printing it doesn't fit in a single page.

Assuming UEs not included in the sample CDR dataset follows similar mobility pattern, the flows (i.e. the number of crossings from one cell to another) for the whole unique LTE users identified in the CDR data is given by

$$\boldsymbol{F_{ij}} = \boldsymbol{H_{ij}} \times \boldsymbol{N_{UE_i}} \qquad (3.5)$$

Figure 3.5 depicts the UE mobility pattern in the LTE network of Addis Ababa, specifically the outflow. The eNodeB sites are represented by Voronoi tessellation of each, and the numbers at the centroid of each cell are set in proportion to the total number of flows from cell i to cell j for UE distribution for the target LTE network. For flow values of 200 and higher (for better visualization), a link is drawn between two cells, and the amount of flow is set in proportion to the thickness of the link.

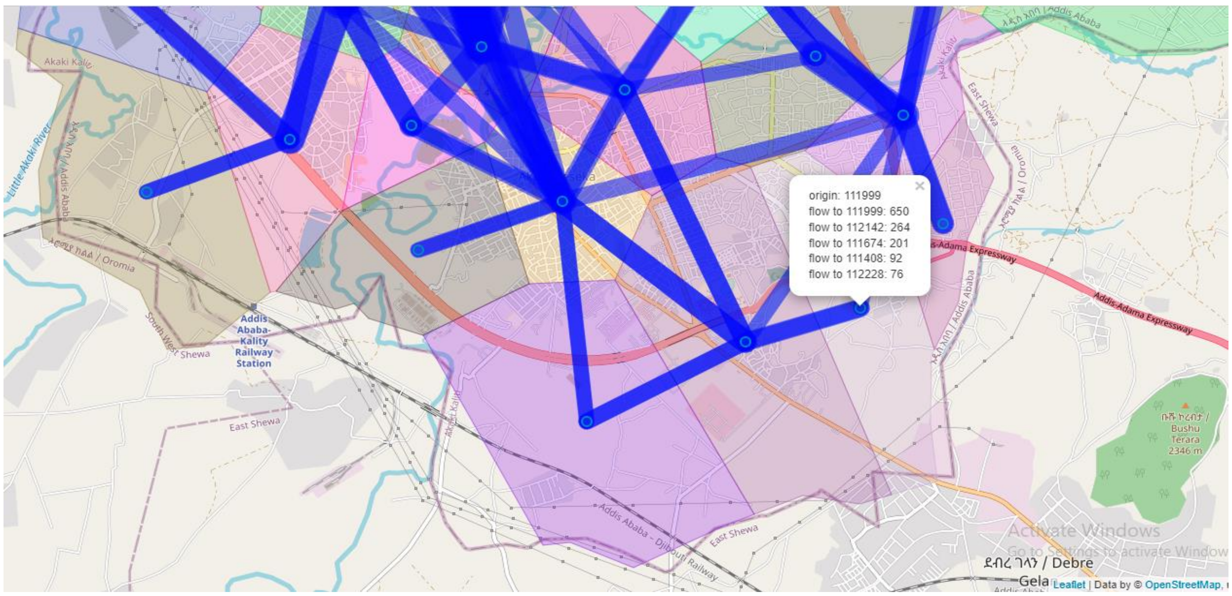


Figure 3. 5 Flow (Crossings) from cell i to cell j

The reconstructed UEs trace data, the flow matrix, and the distribution of UEs in the network will be used for signaling overhead analysis in subsequent chapters. Table 3.3 provides a summary of the mobility data.

Table 3. 3 Summary of Mobility Data

Parameters	**Record length**	**Dimension**	**Reference**
Number of UEs in cell i $\boldsymbol{N_{UE_i}}$	650	650 x 1	
UE traces records ($\boldsymbol{R_C}$)	5,482,544	5482544 x 2	
Number of outflows ($\mathbf{F}_{ij}$)	650	650 x 650	

Chapter 4

Tracking Area Design and Optimization

In this section, the objective function for TA design and optimization of Addis Ababa's LTE network is presented. A heuristic approach, separately embedded with a distance-based clustering algorithm at one time and a density-based clustering algorithm at another time, is used for grouping and re-grouping of cells to produce the best Voronoi distribution of TAs to tackle the overall signaling optimization problem is explained.

4.1. Objective Function

To formulate the overall signaling overhead, let, N_{cell} = {1,..., N} denotes the number of cells in an LTE network, and N_{TA} = {1,..., T} is the number of TAs in the current configuration. Cell-to-TA assignment is represented by the vector **T** = [t_1,..., t_N], where t_i is the TA of cell i. An $N_{cell} \times N_{cell}$ symmetric binary matrix $\mathbf{X_T}$ can be used to indicate TA design **T** where the element $X_{\mathrm{ij}}(\mathbf{T})$ is binary representing whether or not two cells are in the same TA, i.e.,

$$\mathbf{X_T} = \begin{cases} 1 & if\ t_i = t_j \\ 0 & otherwise \end{cases} \quad (4.1)$$

Let $\boldsymbol{N_{UE_i}}$ be the total number of UEs in cell i and the outflow volume, denoted by $\mathbf{F}_{ij}$, be the number of UEs traveling from cell i to cell j over the same period, as defined in chapter three. The signaling overhead of one paging and one tracking area update be symbolized by the letters p_c and t_c, respectively. The precise relationship between t_c and p_c is determined by the number of messages exchanged between the UE and the MME [2]. The call intensity factor/activity factor (∂) is also another parameter (i.e., probability that a UE has to be paged) that should be considered. As tracking area update is initiated by the UE, TAU minimization is critical from the standpoint of UEs because it has a direct impact on power consumption, 10 mw of power is consumed per TAU for the current generation of smartphones and thus battery time [2]. To capture this effect, the contribution to the cost of one TAU (β = 1) to that of a paging procedure a penalizing factor value of $\boldsymbol{\alpha}$ = 10 [9] has been used. The total update and paging signaling overhead $S_{cost}(\mathbf{T})$ defined equation (4.2) is used to formulate the ultimate goal of minimizing the total signaling overhead in the LTE network. The objective function becomes:

$$\underset{\mathbf{T}}{minimize} \quad S_{cost}(\mathbf{T}) = \alpha \times TAU_{cost}(\mathbf{T}) + \beta \times P_{cost}(\mathbf{T}) \tag{4.2}$$

$$subject\ to \qquad N_{TA} \leq N_{cell}, \tag{4.3}$$

$$1 \leq \sum_{i=1}^{N_{cell}} t_{ik} \leq N_{cell}, \forall k, \tag{4.4}$$

$$\sum_{k=1}^{N_{TA}} t_{ik} = 1, \forall i \tag{4.5}$$

The contribution of the TAU, $TAU_{cost}(\mathbf{T})$, to the network signaling overhead is given by:

$$TAU_{cost}(\mathbf{T}) = \sum_{i=1}^{N_{cell}} \sum_{j=1}^{N_{cell}} (\mathbf{1} - X_{ij}) \times F_{ij} \times t_c \tag{4.6}$$

Where,

- N_{cell} is the number of cells in the target LTE network
- X_{ij} are the components of the matrix $\boldsymbol{X_T} \in B^{N_{cell} \times N_{cell}}$ generated from **T.** In particular, $X_{ij} = 1$ if cell i and j belong to the same TAs, and 0 otherwise.
- F_{ij} are the elements of the flow matrix $\mathbf{F} \in \mathrm{N}^{N_{cell} \times N_{cell}}$. It characterizes the number of UEs moving from cell i to cell j during the evaluation time.
- t_c, is the number of messages involved for a single trucking update.

The paging overhead, $P_{cost}(\mathbf{T})$, introduced in cell j while paging UEs in cell i is the second term, and is computed using the equation (4.7) if the two cells are in the same TA.

$$P_{cost}(\mathbf{T}) = \sum_{k=1}^{N_{TA}} ((\partial \times \boldsymbol{N_{UE}}) . \mathbf{X_T} . \sum_{i=1}^{N_{cell}} t_{ik}) \times p_c \tag{4.7}$$

Where,

- $(\partial \times \boldsymbol{N_{UE}}) \in N^{N_{cell}}$ is a vector containing the number of incoming connections in cell i, computed as ∂ multiplied by the number of UEs in each cell.
- Hence, $(\partial \times \boldsymbol{N_{UE}}) \,.\, \mathbf{X_T} \in \boldsymbol{N}^{N_{TA}}$ is a vector containing the aggregation of incoming connections in each cell of each TA.
- $\sum_{i=1}^{N_{cell}} t_{ik} \in N^{N_{TA}}$ is the number of cells in each TA.
- p_c, is the number of messages involved for single paging attempt

4.2. Signaling Overhead Optimization Approach

The goal of the heuristic approach embedded with the clustering algorithms is to find a set of meaningful TAs that minimizes the overall signaling overhead that can be used as a basis for the TAL design and optimization, though TAL optimization is not covered in this thesis. The first clustering technique (K-means) [27] considered is distance-based while the second is density-based clustering technique known as hierarchical density-based spatial clustering application with noises (HDBSCAN) [10].

4.2.1. Assigning Cells to TAs (Kmeans Clustering Approach)

To design the Voronoi distribution of TAs that minimizes the overall signaling overhead, the k-means clustering approach is used. This is a quick approach for generating and comparing possible TA distributions, as it only takes Euclidian distance to group cells. Algorithm 1a, adapted from [9], depicts the entire procedure, with specific lines referred to in the following explanation.

The algorithm takes n, which is the possible number of clusters and the cell coordinates ($\boldsymbol{x}$) in the UE's traces and intelligently initializes N_{TA}^* initial means, and creates N_{TA}^* TAs based on the Voronoi cell geometry. The TA centroids (new means, $\boldsymbol{\mu}$) are computed from cell coordinates once all of the cells have been allocated to a TA. Subsequently, each cell is reassigned to the TA with the shortest squared Euclidean distance between its centroid and the TA centroid. The procedure recurrently updates the Voronoi distribution till it converges to the smallest sum squared distance feasible. The program searches through all of the potential TA numbers (line 1).

Algorithm 1a: K-means clustering for allocation of cells to TAs

Input:

$\{\beta, \alpha, \partial, p_c, t_c, N_{cell}\}$

$\boldsymbol{x} \in \mathbb{R}^{N_{cell} \times 2}$: coordinates of all cells

$\boldsymbol{N_{UE}} \in \mathbb{N}^{N_{cell} \times 1}$: number of UEs in each cell

$\mathbf{F} \in \mathrm{N}^{N_{cell} \times N_{cell}}$: the number of UEs moving from cell i to cell j

Output:

$\{\mathbf{T}^*, S^*, N_{TA}^*\}$

1 **for** 1 N_{cell} do

2 // Smartly guess n number of centroids
$\boldsymbol{\mu} \leftarrow kmeans(\boldsymbol{x}, n)$

3 //Get cluster labels
$\{\boldsymbol{T'}\} \leftarrow kmeans(\boldsymbol{x}, n, \boldsymbol{\mu})$

4 //Get unique number of labels
$N_{TA}^* = count(unique\ cluster\ labels\ in\ \boldsymbol{T'})$

5 $\boldsymbol{X_T} \leftarrow Evaluate\ same\ TA\ membership\ (\boldsymbol{T'})$

6 $TAU_{cost} \leftarrow Compute\ TAU\ cost((\mathbf{1} - \boldsymbol{X_T}), \mathbf{F}, t_c)$

7 $P_{cost} \leftarrow Compute\ Paging\ cost(\boldsymbol{X_T}, \partial, \boldsymbol{N_{UE}}, p_c)$

8 $S_{cost} \leftarrow Compute\ total\ signaling\ cost(TAU_{cost}, P_{cost}, \beta, \alpha,)$

9 **if** n=1 **then**

10 //Compute the baseline signaling cost
$\{\mathbf{T^b}, S^b, N_{TA}^b\}$

11 $\{\mathbf{T}', S_{cost}, n\}$ //Compute the new signaling cost

12 //compare and select the minimum of the two
$\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \min(\{\mathbf{T^b}, S^b, N_{TA}^b\}, \{\mathbf{T}', S_{cost}, n\})$

13 **else if** $S_{cost} < S^*$ **then**

14 $\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \{\mathbf{T}', S_{cost}, n\}$

15 **end**

The first N_{TA}^{*} centroids ($\boldsymbol{\mu}$) (line 2) are chosen by k-means and it returns the Voronoi distribution labels (**T**) with the smallest feasible sum squared distance (line 3). An objective function reflecting the signaling overhead in the entire network is computed for each TA setup. Given this, eq. (4.6) is used to calculate the contribution of TAUs to network signaling overhead (lines 5-6). Paging overhead, on the other hand, is given by (line 7) and is calculated using eq. (4.7). Equation (4.2) is used to calculate the overall signaling overhead for each TA set up in the network (line 8). To get a reference signaling overhead, for improvement evaluation, the signaling cost for the baseline TA configuration is computed once (line 10). Finally, the algorithm selects the N_{TA}^{*} and $\mathbf{T}^{*}$ matrices with the lowest signaling cost, S^{*}. (lines 9-14).

4.2.2. Assigning Cells to TAs (HDBSCAN Clustering Approach)

In this approach, HDBSCAN [10] is used to design the Voronoi distribution of TAs that minimizes overall signaling overhead. HDBSCAN is a hierarchical clustering algorithm that transforms the space based on density/sparsity, constructs a cluster hierarchy of connected components, condenses the cluster hierarchy based on minimum cluster size, and extracts the stable clusters from the condensed tree to produce flat clustering. It computes the stability for each cluster as $\sum_{x \in cluster}(\lambda_x - \lambda_{birth})$, where $\lambda = \frac{1}{distance}$, and λ_{birth} is the value when the cluster splits off and becomes its own cluster. If the sum of the child clusters' stabilities is greater than the cluster's stability, the cluster's stability is set to be the sum of the child stabilities. If, on the other hand, the stability of the cluster exceeds the sum of its children, it asserts the cluster to be chosen and unselects all of its descendants. The current set of selected clusters is flat clustering once it reaches the root node. It has the λ_x for each point $\boldsymbol{x}$ in each cluster and normalizes those values (so they range from zero to one) to have a measure of the strength of cluster membership for each point in the cluster and returns the clusters with labels [10]. The entire procedure is depicted in Algorithm 1b, adapted from [9], with specific lines referred to in the following explanation. The algorithm takes the minimum number of points (density, k) - the smallest size grouping to be considered a cluster, the minimum number of samples to be considered for distance computation, the cluster selection epsilon which guarantees that clusters below the given threshold are not split up any further, and the set of routes of all UEs ($\boldsymbol{R_C}$) with the cell coordinates ($\boldsymbol{x}$). The program searches through all of the potential density numbers (k) (line 1). The computes the stability and returns the Voronoi distribution labels (**T**) based on the probability of a cell's membership to a TA (line 3). An objective function reflecting the signaling overhead in the entire network is computed for each TA setup.

Algorithm 1b: HDBSCAN clustering for allocation of cells to TAs

Input:

$\{\beta, \alpha, \partial, p_c, t_c\}$

$\boldsymbol{x} \in \mathbb{R}^{N_{cell} \times 2}$: coordinates of all cells

$\boldsymbol{R_C} \in \mathbb{R}^{N_{nodes} \times 2}$: the set of routes with coordinates of the cells for all UEs

$\boldsymbol{N_{UE}} \in \mathbb{N}^{N_{cell} \times 1}$: number of UEs in each cell

$\mathbf{F} \in \mathbb{N}^{N_{cell} \times N_{cell}}$: the number of UEs moving from cell i to cell j

$k - the\ minimum\ size\ grouping\ to\ consider\ as\ a\ cluster - density$

$K \in \mathbb{N} - list\ of\ k - variable$

$s = 1{:}\ - minimum\ number\ of\ samples$

$\varepsilon = 700\ m\ - \ the\ distances\ between\ cells\ coordinate$
$to\ stop\ separating\ clusters\ that\ are\ less\ than\ \varepsilon$

Output:

$\{\mathbf{T}^*, S^*, N_{TA}^*\}$

1 **for** k in K do

2 //Initialize iteration number

$i = 0$

3 //Get cluster labels

$\{\boldsymbol{T}'\ \} \leftarrow HDBSCAN(\boldsymbol{R_C}, k, s, \varepsilon)$

4 //Get the unique number of labels

$N_{TA}^* = count(unique\ cluster\ labels\ in\ \boldsymbol{T}')$

5 $\boldsymbol{X_T} \leftarrow Evaluate\ same\ TA\ membership\ (\boldsymbol{T}')$

6 $TAU_{cost} \leftarrow Compute\ TAU\ cost((\mathbf{1} - \boldsymbol{X_T}), \mathbf{F},\ t_c)$

7 $P_{cost} \leftarrow Compute\ Paging\ cost(\boldsymbol{X_T}, \partial, \boldsymbol{N_{UE}}, p_c)$

8 $S_{cost} \leftarrow Compute\ total\ signaling\ cost(TAU_{cost}, P_{cost}, \beta, \alpha,)$

9 **if** n=1 **then**

10 //Compute the baseline signaling cost

$\{\mathbf{T^b}, S^b, N_{TA}^b\}$

11 $\{\mathbf{T}', S_{cost}, n\}$ //compute the new signaling cost

12 //compare and select the minimum of the two

$\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \min(\{\mathbf{T^b}, S^b, N_{TA}^b\}, \{\mathbf{T}', S_{cost}, n\})$

13 **else if** $S_{cost} < S^*$ **then**

14 $\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \{\mathbf{T}', S_{cost}, n\}$

15 $i = i + 1$ // increment the iteration number by one

16 **end**

Given this, eq. (4.6) is used to calculate the contribution of TAUs to network signaling overhead (lines 5-6). Paging overhead, on the other hand, is given by (line 7) and is calculated using eq. (4.7). Equation (4.2) is used to evaluate the overall signaling overhead for each TA set up in the network (line 8) for each set of TAs in each iteration. To get a reference signaling overhead, for improvement evaluation, the signaling cost for the baseline TA configuration is computed once (line 10). Finally, the algorithm selects the N_{TA}^{*} and $\mathbf{T}^{*}$ matrices with the lowest signaling cost, S^{*}. (lines 9-14).

4.3. Performance Evaluation

The achieved minimum overall signaling overhead in terms of the number of messages, from each clustering technique is evaluated against the baseline signaling overhead, for the same UE trace, UE distribution dataset and same LTE network topology. The minimum signaling overhead achieved by each iteration for each approach is measured in terms of the total number of messages exchanged between the UEs and the MME, and is compared against the baseline.

Let, $S_{cost}(\mathbf{T}^{b})$ be the total signaling overhead for the baseline configuration and $S_{cost}(\mathbf{T}^{*})$ be the total signaling overhead for the new TAs, the performance improvement is evaluated using the equation (4.8)

$$Improvement\ (\%) = \frac{S_{cost}(\mathbf{T}^{b}) - S_{cost}(\mathbf{T}^{*})}{S_{cost}(\mathbf{T}^{b})} \times 100\ \% \qquad (4.8)$$

Chapter 5

Results and Discussions

This chapter presents, the results and discussion. The specific parameters and values used for the signaling overhead calculation as well as the results and performance evaluation for the distance-based clustering approach and the density-based clustering approach, against the baseline overall signaling overhead for the same scenario (i.e. network topology, UE distribution, and mobility pattern), in the LTE network of Addis Ababa is presented.

5.1. Parameters and Values

The parameters and values used for signaling overhead computation, in this thesis, are shown in Table 5.1. The level of paging volume is defined using parameter ∂ = {0.2} which indicates the percentage of UEs being paged [9]. That would represent an average paging rate of 20 pages/h per subscriber. This value is in line with the data considered in other investigations [9]. It is important to note that with the widespread use of smartphones and push messages, paging volumes have increased significantly [7], [2], [9]. As tracking area update is initiated by the UE, TAU minimization is critical from the standpoint of UEs because it has a direct impact on power consumption and thus battery time. To capture this effect, the contribution to the cost of one TAU (β = 1) to that of a paging procedure a penalizing factor value of $\boldsymbol{\alpha}$ = 10 [9] has been used. Another parameter ($t_c = 40$), is used as tracking area update cost to account for the number of messages exchanged for reporting a change in location by the UE, which amounts to 40 messages per TAU. Similarly, when the MME initiates the paging process for delivering UE-specific data packets, the total number of messages exchanged is 33 messages per paging, so the parameter cost of paging ($p_c = 33$) [2].

Table 5. 1 Parameters and Values

Parameters	Values	Remarks
Number of Cells (N_{cell})	650	Number eNodeBs
Baseline number of TAs (N_{TA}^{b})	8	Number of TAs for the current configuration
Paging rate ∂	20 %	Assumed, due to the data usage and push notification [9]
Paging factor β	1	
Tracking update factor $\boldsymbol{\alpha}$	10	TAU to paging cost (10:1), due to UE battery utilization and battery life [9]
Paging cost, p_c	33 messages	33 messages for one paging attempt [2]
Tracking update cost t_c	40 messages	40 messages for one tracking area update [2]

5.2. Results

The minimization objective function defined in chapter 4 and the cell to TAs assignment algorithms explained in chapter 4 is run in a python environment having the distance-based k-means clustering library and the density-based HDBSCAN clustering library, on a machine with Windows 10 operating system and 8 GB memory (RAM). The results of the computations for each of the approaches are explained subsequently.

5.2.1. Signaling Overhead (Baseline)

The location management signaling overhead, for the baseline TAs (8) configuration in the LTE network understudy, is found to be $S_{cost}(\mathbf{T}) = 6.08e + 08$ (Table 5.2). This signaling overhead is used as the baseline for performance evaluation for the minimum overall signaling overhead calculated based on the K-means clustering approach and the density-based HDBSCAN clustering approach.

Table 5. 2 Signaling Overhead (Baseline Configuration, N_{TA} = 8)

Algorithm	Number of TAs	Tracking area update cost $TAU_{cost}(\boldsymbol{T})$	Paging cost $P_{cost}(\boldsymbol{T})$	Total signaling cost $S_{cost}(\boldsymbol{T})$	Improvement (%) $\frac{S_{cost}(\boldsymbol{T}^b) - S_{cost}(\boldsymbol{T}^*)}{S_{cost}(\boldsymbol{T}^b)}$
Baseline	8	1.30e+08	4.78e+08	6.08e+08	

5.2.2. Signaling Overhead (K-means Clustering Approach)

The location management signaling overhead trend for the K-means clustering approach is shown in Figure 5.1. from which, one can observe that the worst case in terms of the signaling cost is when we consider the allocation of 650 cells into one TA, i.e., N_{TA} = 1. Due to the contribution of the P_{cost}, we also observe that the signaling cost decreases up to N_{TA} = 74, which is the optimum number of TAs found. After that, it increases back again, in this, case due to the contribution of TAU_{cost}.

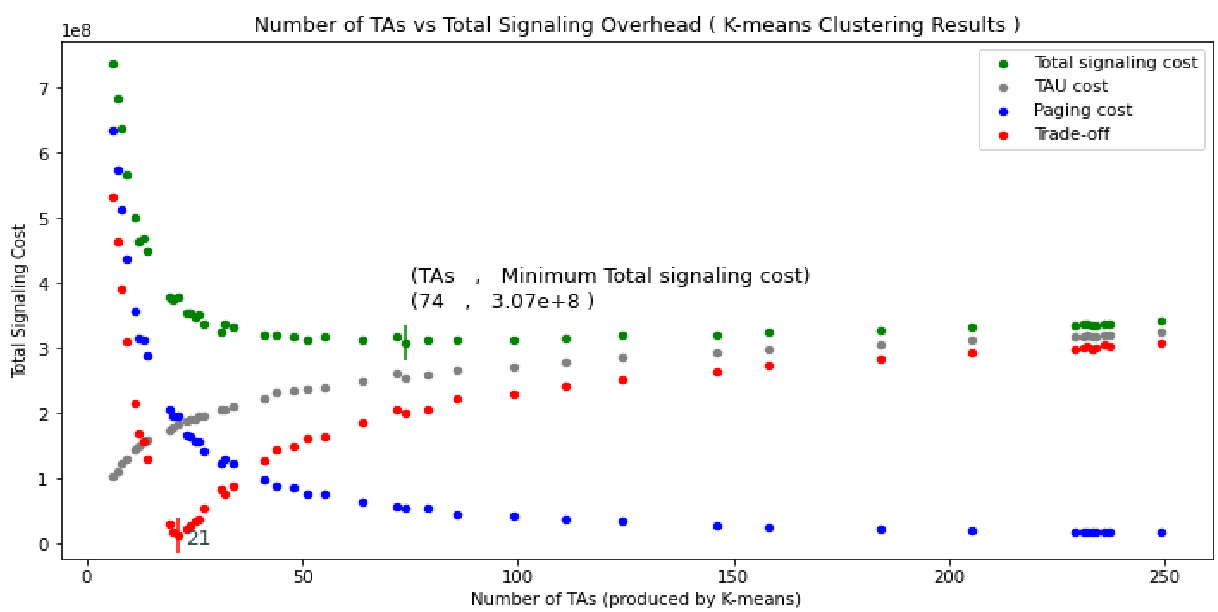


Figure 5. 1 Signaling overhead (K-means clustering approach) for each of the 43 solutions.

The cell distribution (eNodeB level) for each TA is shown in Figure 5.2. The cell distribution, for example, with the absolute difference between the tracking update signaling load and paging load is at minimum has 21 TAs (N_{TA} = 21) which results in an overall signaling load of $S_{cost}(\mathbf{T}) = 3.79\mathrm{e}+08$, having a performance improvement of 37.73 % as compared to the baseline TA configuration. Likewise, the result that relates to the optimum number of TAs (74) where the overall signaling overhead is found to be minimum has a total signaling overhead of $S_{cost}(\mathbf{T}) = 3.07\mathrm{e}+08$ which has an improvement of 49.45 % as compared to the baseline TA configuration.

(a) Baseline Clusters, (Number of TAs = 8)

(b) K-means Clusters for minimum trade-Off (Number of TAs =21)

(c) K-means Clusters for minimum total signaling overhead (Number of TAs =74)

Figure 5. 2 Cell distributions to TAs (K-means clustering approach).

The cell distribution, having 650 TAs (N_{TA} = 650) has an overall signaling load of $S_{cost}(\mathbf{T}) = 3.77\mathrm{E}+08$, giving a performance improvement of 37.98 % as compared to the baseline TA configuration. However, the cell distribution, having 1 TAs (N_{TA} = 1), i.e. each cell is assigned to be a TA, has an overall signaling overhead of $S_{cost}(\mathbf{T}) = 3.96\mathrm{E}+09$, produces a degradation of performance by 551.72 % as compared to the baseline TA configuration. The summary of the results of the k-means approach, is shown in Table 5.3.

Table 5. 3 Signaling Overhead (K-means)

Algorithm	Number of TAs	Tracking area update cost $TAU_{cost}(\boldsymbol{T})$	Paging cost $P_{cost}(\boldsymbol{T})$	Total signaling cost $S_{cost}(\boldsymbol{T})$	Improvement (%) $\frac{S_{cost}(\boldsymbol{T}^b) - S_{cost}(\boldsymbol{T}^*)}{S_{cost}(\boldsymbol{T}^b)}$
One Cluster	1	0.00E+00	3.96E+09	3.96E+09	-551.72 %
Baseline	8	1.30e+08	4.78e+08	6.08e+08	
K-means*	21	1.83e+08	1.96e+08	3.79e+08	37.73 %
K-means**	74	2.54e+08	5.31e+07	3.07e+08	49.45 %
Single Cell TAs	650	3.71E+08	6.10E+06	3.77E+08	37.98 %

The signaling overhead as a result of the HDBSCAN clustering approach is presented next.

5.2.3. Signaling Overhead (HDBSCAN Clustering Approach)

The location management signaling overhead for the K-means clustering approach is shown in Figure 5.3. From which, one can observe that the worst case in terms of the signaling cost is when we consider the allocation of 650 cells into six TAs, i.e., N_{TA} = 6 which is the minimum number of TAs produced by HDBSCAN and the maximum number of TAs that produces and keeps the total overall signaling load at minimum are N_{TA}= 237. In line with the decrease of the contribution of the paging cost, $P_{cost}(\boldsymbol{T})$, one can also detect that the total signaling cost decreases up to N_{TA}= 237 and stabilizes then after, as HDBSCAN produces the same clusters, TAs having the cell distribution as shown in Figure 5.4. The cell distribution, for example, with the absolute difference between the tracking update signaling load and paging load is at minimum has 99 TAs (N_{TA} = 99) which results in an overall signaling load of $S_{cost}(\mathbf{T}) = 4.47\text{E} + 08$, amounting to a performance improvement of 26.45 % as compared to the baseline TA configuration. Likewise, the result that relates to the optimum number of TAs (N_{TA} = 237) where the overall signaling overhead is found to be minimum has a total signaling overhead of $S_{cost}(\mathbf{T}) = 2.46\text{e} + 08$ which has an improvement of 59.54 % as compared to the baseline TA configuration.

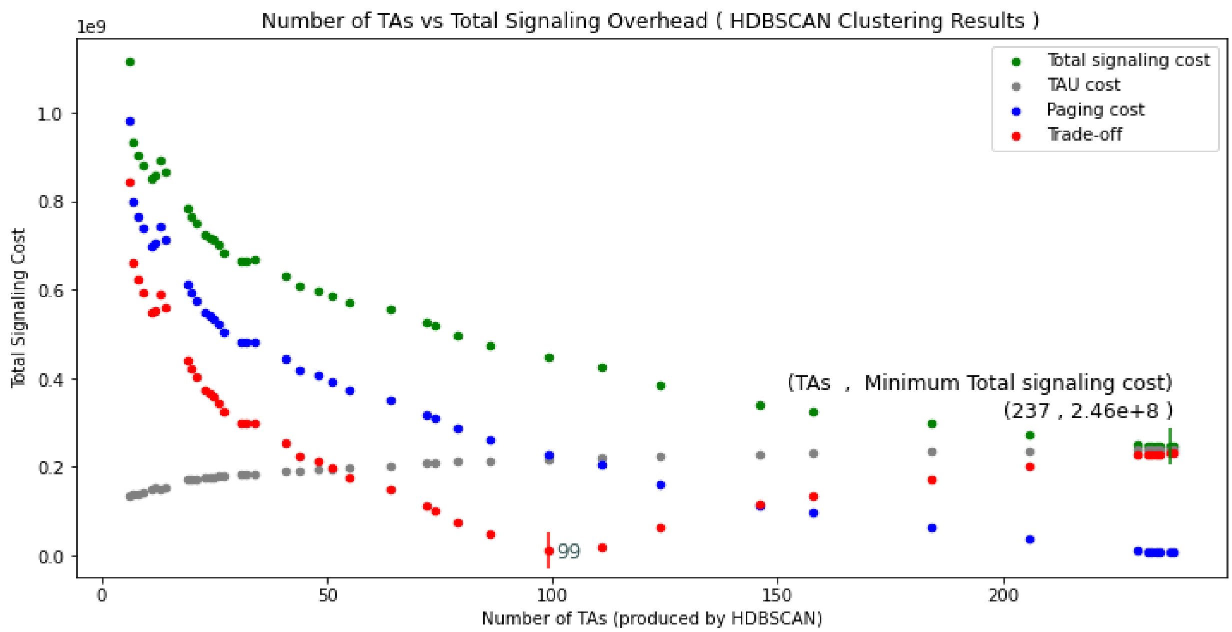
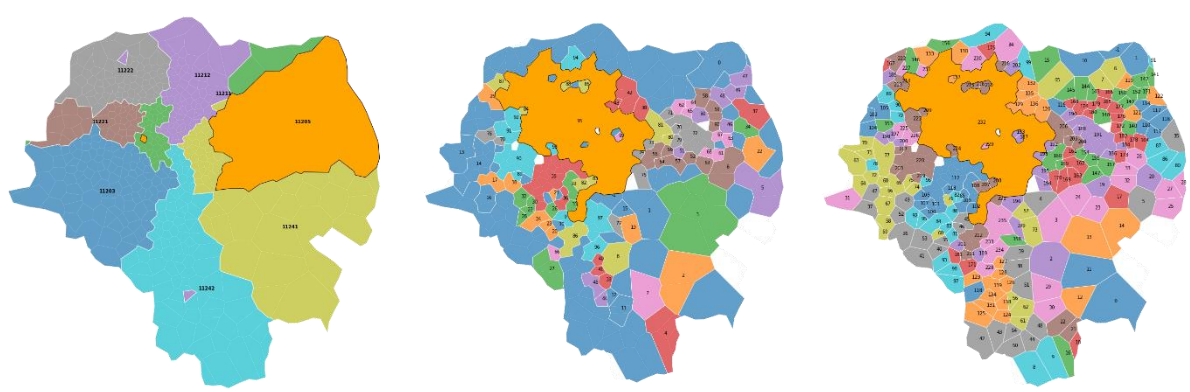


Figure 5. 3 Signaling overhead of each of the 43 unique solutions (HDBSCAN).



(a) Baseline Clusters, (Number of TAs = 8)

(b) HDBSCAN Clusters for minimum trade-Off (Number of TAs =99)

(c) HDBSCAN Clusters for minimum total signaling overhead (Number of TAs = 237)

Figure 5. 4 Cell distribution (HDBSCAN clustering approach).

The summary of the results of the HDBSCAN clustering approach, is shown in Table 5.4 below.

Table 5. 4 Signaling Overhead (HDBSCAN)

Algorithm	Density	Number of TAs	Tracking area update cost $TAU_{cost}(\boldsymbol{T})$	Paging cost $P_{cost}(\boldsymbol{T})$	Total signaling cost $S_{cost}(\boldsymbol{T})$	Improvement (%) $\frac{S_{cost}(\boldsymbol{T}^b) - S_{cost}(\boldsymbol{T}^*)}{S_{cost}(\boldsymbol{T}^b)}$
Baseline		8	1.30E+08	4.78E+08	6.08E+08	
HDBSCAN*	8000	99	2.17E+08	2.30E+08	4.47E+08	26.45 %
HDBSCAN**	400	237	2.38E+08	7.79E+06	2.46E+08	59.54 %

The signaling overhead performance evaluation for all of the approaches is presented in the next chapter.

5.3. Discussions

The location management in the LTE network of Addis Ababa is found to be the standard scheme that uses TA to concurrently broadcast paging messages to all cells in a TA and requires a UE to send TAU every time it crosses a TA border in the network. The deployed TA plan is based on a site-level coverage area. In other words, an eNodeB site along with the cells configured in it is configured to a single TA. It is also found out that 650 eNodeBs (2206 cells) are distributed to 8 TAs as presented in Table 2.4.

The CDR data of UEs can be used to rebuild a good approximation of UE's traces within the LTE network. For example, the 574,610 rows of CDR data of the 9307 LTE users studied in this thesis is rebuilt through the shortest path analysis on the drivable road network of the city, generating 5,482,544 rows of UE traces, which in turn is used for the mobility analysis.

From the mobility pattern perspective, some cells are identified to have relatively high crossings or flow to other cells while others have low. For example, out of the 5044 UEs identified to exist at cell '111577' the number of UEs moving to cell '111001' is found out to be 1631, though the number of crossing or flow from '111577' to other cells is zero. The distribution of UEs and the flow values is already presented in the mobility analysis section of this thesis, chapter 3. This result suggests that the TA design has to study the density and mobility pattern of the LTE network users of Addis Ababa.

Table 5. 5 Signaling Overhead Performance Evaluation of all approaches

Algorithm	Density	Number of	Tracking area update cost $TAU_{cost}(\boldsymbol{T})$		Paging cost $P_{cost}(\boldsymbol{T})$		Total signaling cost $S_{cost}(\boldsymbol{T})$	Improvement (%) $\frac{S_{cost}(\boldsymbol{T}^b) - S_{cost}(\boldsymbol{T}^*)}{S_{cost}(\boldsymbol{T}^b)}$
			Num. of messages	(%)	Num. of messages	(%)	Num. of messages	(%)
Baseline		8	1.30E+08		4.78E+08		6.08E+08	
One Cluster		1	0.00E+00	**100 %**	3.96E+09	**-728 %**	3.96E+09	**-551.32 %**
K-means		74	2.54E+08	**-95 %**	5.31E+07	**89 %**	3.07E+08	**49.51 %**
HDBSCAN	400	237	2.38E+08	**-83 %**	7.79E+06	**98 %**	2.46E+08	**59.54 %**

The signaling overhead performance evaluation of all the approaches is presented in Table 5.5 above and the following discussion is based on that. The overall signaling overhead of the baseline TAs (8) configuration is found to be $S_{cost}(\mathbf{T}) = 6.08\text{E} + 08$ which has a TAU component of 1.30E+08 messages and a paging component of 4.78E+08.

The results for TAs created with K-means clustering approach shows that the worst case in terms of the signaling cost is when the allocation of 650 cells into one TA is done, i.e., N_{TA} = 1, producing 551.32 % more overall signaling overhead as compared to the baseline overall signaling overhead, as the contribution of the P_{cost} is 100 %, even though the TAU component is effectively zero. Creating TAs having a single cell as a member, i.e., N_{TA} = 650, has 99 % improvement for paging overhead but produces 185 % more tracking area update overhead, which results in a 37.99 % overall signaling overhead improvement as compared to the baseline. The grouping of the cells into 74 TAs (N_{TA} = 74) shows that the tracking area update overhead is 95 % more than the baseline TAU overhead while the paging overhead is 89 % less than the corresponding baseline paging signaling overhead, though the total overall minimum signaling cost $S_{cost}(\mathbf{T}) = 3.07\text{E} + 08$ has 49.45 % improvement as compared to the baseline overall signaling overhead.

Similarly, the results for TAs created with HDBSCAN clustering approach (N_{TA} = 237) shows that the TAU overhead is 83 % more than that of the baseline TAU overhead while the paging overhead is 98 % less than that of the baseline paging overhead, though the total overall minimum signaling cost $S_{cost}(\mathbf{T}) = 2.46\mathrm{E}+08$ has 59.54 % improvement as compared to the baseline overall signaling overhead. Therefore, TAs created with HDBSCAN clustering approach has a better performance (59.54 %) in terms of the total signaling overhead against the baseline TAs, due to the fact that HDBSCAN considers the density of the LTE users while grouping cells into TAs. Figure 5.5 shows the total signaling overhead trend vs the number of TAs (aligned with the same N_{TA}): for the baseline, for TAs created with K-means clustering approach and for TAs created with HDBSCAN clustering approach.

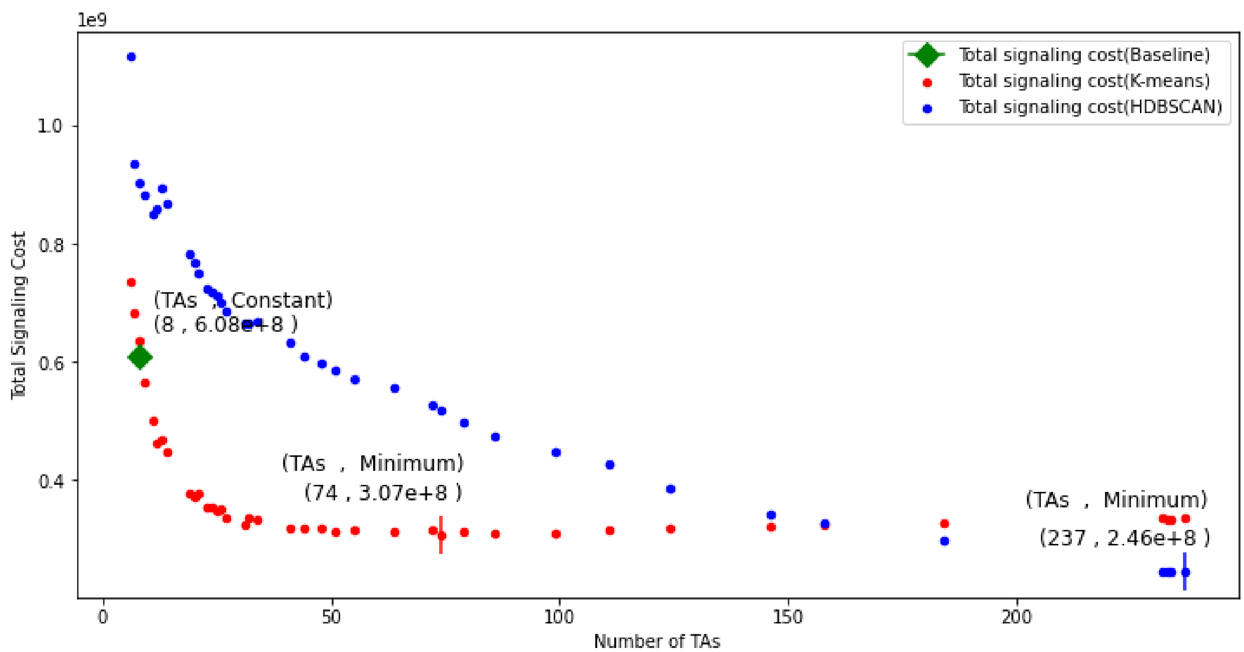
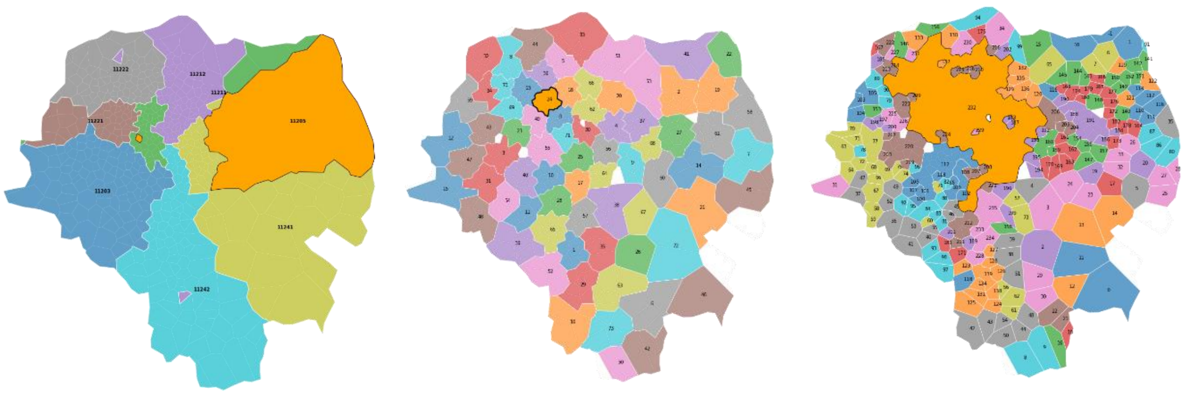


Figure 5. 5 Total signaling overhead trend (aligned with the same N_{TA}).



(b) Baseline Clusters, (Number of TAs = 8)

(b) K-means Clusters for minimum total signaling overhead (Number of TAs =74)

(c) HDBSCAN Clusters for minimum total signaling overhead (Number of TAs = 237)

Figure 5. 6 Cell distribution (for the minimum total overall signaling overhead).

Figure 5.6 shows the cell distribution for the minimum total signaling overhead for the baseline, for TAs created with K-means clustering approach and for TAs created with HDBSCAN clustering approach.

Chapter 6

Conclusions and Recommendations

Following the TA design and optimization in LTE networks analysis in Chapter 5, the conclusions and recommendations are presented in this chapter.

6.1. Conclusions

The tracking area design and optimization approach for location management signaling overhead in LTE networks is confirmed to adjust to changes and trends of UE location and mobility patterns, though the sample data used for analysis may have some bias. For real networks having a large number of cells, creating TAs with single-cell member-only or assigning all cells to a single TA, or keeping the TAs deployed during initial stages may not be optimum for signaling overhead. To tackle the problem of allocating cells to TAs in such a way that the overall signaling overhead is minimized a heuristic algorithm, independently and separately embedded with two clustering techniques, has been implemented. In the first heuristic algorithm, the K-means clustering technique is embedded within it and produces relatively good results, as compared to the baseline. In the second heuristic algorithm, the HDBSCAN clustering technique is embedded within it and results are better than the baseline as well as the K-means clustering technique, as HDBSCAN takes the density and mobility pattern of the users into account to assign cells to TAs. The case study on a realistic network using the standard location management schemes shows that the heuristic algorithm embedded with the HDBSCAN clustering technique is able to improve the signaling overhead significantly.

6.2. Recommendation

As the sampling technique employed is random (i.e. user profiling is not used), the performance results of the analysis may have some bias. Taking the bias in mind, the following recommendations are made:

- The case study network location management signaling overhead can be improved significantly with the use and implementation of the TA design and optimization approach presented in this thesis.
- For efficient dynamic TAL implementation, the TA design and optimization approach presented in this thesis can be used as a foundation.
- The work in this thesis can be extended by considering algorithms that assign TA into TAL and implement TAL assignment to individual UEs in the LTE network through the incorporation of individual mobility patterns and incoming call connection characteristics.

References

[1] Ericsson, "Ericsson Mobility Report," SE-164 80 Stockholm, Sweden, 2021.

[2] Alaa A. R. Alsaeedy and Edwin K. P. Chong, "A review of mobility management entity in LTE networks:," *Int J Network Mgmt,* no. e2088, 2019.

[3] 3GPP TS 23.401 v14.4.0., *General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access (release 14),* 2017.

[4] Sara Modarres Razavi, *Tracking Area Planning in Cellular Networks - Optimization and Performance Evaluation,* Sweden: Link¨ oping Studies in Science and Technology, 2011.

[5] Razavi SM, Yuan D, Gunnarsson F and Moe J, *Exploiting tracking area list for improving signaling overhead in LTE,* 2010, pp. 1-5.

[6] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu and Wenye Wang, "Mobility management in next-generation wireless systems," *in Proceedings of the IEEE,* vol. 87, no. 8, pp. 1347-1384, Aug 1999.

[7] Aqeeli E, Hashim HA, Haque A and Shami A, "Optimal location management in LTE networks using evolutionary techniques," *Int J Commun Syst,* vol. e4100, 2019.

[8] E. Aqeeli, A. Moubayed and A. Shami, "Dynamic SON-Enabled Location Management in LTE Networks," *in IEEE Transactions on Mobile Computing,* vol. 17, no. 7, pp. 1511-1523, 1 Jul 2018.

[9] Jessica Moysen and Mario Garc´ıa-Lozano, "Learning-based Tracking Area List Management in 4G and 5G Networks," *in IEEE Transactions on Mobile Computing,* vol. 19, no. 8, pp. 1862-1878, 1 Aug 2020.

[10] Github, "Understanding HDBSCAN and Density-Based Clustering," 17 Jan 2021. [Online]. Available: https://pberba.github.io/stats/2020/01/17/hdbscan/. [Accessed 15 Dec 2021].

[11] Emad Aqeeli, Abdallah Moubayed and Abdallah Shami, "Towards Intelligent LTE Mobility Management through MME Pooling," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015.

[12] Kuo Hsing Chiang and Nirmala Shenoy, "A 2D Random Walk Mobility Model for Location Management Studies in Wireless Networks," *IEEE Trans Veh Technol,* vol. 53, no. 2, p. 413–424, 2004.

[13] Modarres Razavi S and Yuan D, "Mitigating mobility signaling congestion in LTE by overlapping tracking area lists," in *Proceedings of the 14th acm international conference on modeling, analysis and simulation of wireless and mobile systems*, New York, 2011.

[14] Sara Modarres Razavi, Di Yuan, Fredrik Gunnarsson and Johan Moe, "Dynamic Tracking Area List configuration and performance evaluation in LTE," *2010 IEEE Globecom Workshops,* pp. 49-53, 2010.

[15] Ethiotelecom, *Addis Ababa's LTE network plan,* Addis Ababa: Ethiotelecom, 2021.

[16] I. Demirkol, C. Ersoy, M.U. Caglayan and H. Delic, "Location area planning in cellular networks using simulated annealing," *IEEE Conference on Computer Communications,* vol. INFOCOM, no. 01, pp. 13-20, 2001.

[17] Daqing Gu and Stephen S. Rappaport, "Mobile user registration in cellular systems with overlapping location areas," *1999 IEEE 49th Vehicular Technology Conference (Cat. No.99CH36363),* vol. 1, pp. 802-806, 1999.

[18] T. Winter, "Mobility management and network design for UMTS," in *Proc. of IEEE International Symposium on Personal, Indoor and*, 2004.

[19] GeoPandas, "The documentation of GeoPandas," The documentation of GeoPandas, [Online]. Available: https://geopandas.org/en/stable/. [Accessed 1 November 2021].

[20] Humdata (HDX), "Ethiopia - Subnational Administrative Boundaries," November 2021. [Online]. Available: https://data.humdata.org/dataset/ethiopia-cod-ab. [Accessed 01 November 2021].

[21] OpenStreetMap, "Open Street Map," OpenStreetMap, 1 November 2021. [Online]. Available: https://www.openstreetmap.org/#map=13/9.0174/38.8372. [Accessed 1 November 2021].

[22] OSMnX, "User reference for the OSMnx package.," OSMnX, [Online]. Available: https://osmnx.readthedocs.io/en/stable/osmnx.html. [Accessed 1 November 2021].

[23] NetworkX, "Networkx Algorithms Shortest Paths Weighted Bidirectional Dijkstra," NetworkX, [Online]. Available: https://networkx.org/documentation/networkx-

1.8/reference/generated/networkx.algorithms.shortest_paths.weighted.bidirectional_dijkstra.html. [Accessed 1 November 2021].

[24] Scikit-mobility, "Scikit-mobility," Scikit-mobility, [Online]. Available: https://scikit-mobility.github.io/scikit-mobility/. [Accessed 1 November 2021].

[25] Pappalardo, L., Simini and F, "Data-driven generation of spatio-temporal routines in human mobility," *Data Min Knowl,* vol. Disc 32, p. 787–829, 2018.

[26] Chaoming Song, Tal Koren, Pu Wang and Albert-László Barabási, "Modelling the scaling properties of human mobility," *Nature physics,* vol. 6, 13 September 2010.

[27] towardsdatascience, "Understanding K-means Clustering in Machine Learning," 13 Sep 2018. [Online]. Available: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1. [Accessed 15 Dec 2021].

[28] J. O. Abolade, O. A. Fakolujo and A. Orimogunje, "Handover in Mobile Wireless Communication Network - A Review," *Advanced Engineering, management and Science,* vol. 3, no. 9, p. 934–940, 2017.

[29] Dahlman E, Parkvall S and Sköld J, 4G: LTE/LTE-advanced for Mobile Broadband, Second ed., Chennai: Elsevier/Academic Press, 2011.

[30] J. Cowling, *Dynamic location management in heterogeneous cellular networks,* University of Sydney, 2004.

Appendix A

DESIGN OPTIMIZATION OF TRACKING AREA IN LTE NETWORKS USING CLUSTERING TECHNIQUES

The case of Addis Ababa, Ethiopia.

Zelalem Tesfaye

TELECOMMUNICATION
ENGINEERING GRADUATE
PROGRAM
SCHOOL OF ELECTRICAL
AND COMPUTER ENGINEERING
ADDIS ABABA INSTITUTE OF
TECHNOLOGY

ADDIS ABABA, ETHIOPIA
zelalem.tesfayezt@gmail.com

Abstract— The intense competition in the telecommunications market forces operators and service providers to require optimized and cost-effective networks. In the study of Long Term Evolution (LTE) network location management, one of the major challenges is tracing users at a low cost. The location management signaling overhead of Tracking Area Update (TAU) versus paging is a documented performance concern in LTE networks. To respond to dynamic changes in UE distribution and UE mobility patterns, TA design must be modified over time, because the initial deployment may not be efficient in terms of signaling overhead. This thesis is focused on the planning and optimization of tracking area configurations by taking a real LTE network topology in the city of Addis Ababa, Ethiopia. Two heuristic planning and optimization approaches are implemented to assign and reassign cells to new TAs to find out TAs that improve the overall signaling overhead in the network studied.

The baseline configuration has 650 sites grouped into 8 TAs having an overall signaling overhead of 6.08E+08 messages. The first heuristic algorithm, which embeds the K-means clustering approach, has created 74 TAs with a 49.51 % improvement in total signaling overhead

as compared to the baseline configuration. The total signaling overhead improvement achieved by the second heuristic algorithm using the Hierarchical Density-Based Spatial Clustering of Application with Noises (HDBSCAN) clustering approach is found to be 59.54 %, with 237 TAs created by it.

Keywords: Signaling Overhead Optimization, MME, TA, TAL, LTE, TAU, HDBSCAN

I. Introduction

In the last few decades, the field of wireless and mobile communications has seen marvelous expansion in both capacity and coverage. Nearly 95 % of the world's population is covered with the 3rd Generation Partnership Project (3GPP) cellular technologies, one of them is Long-Term Evolution (LTE) having a projected 4.7 billion subscriptions by the end of 2021. There are more than 5.5 billion new smartphone subscriptions supporting LTE communication [1]. Despite the rapid growth, LTE networks continue to be challenged by severe signaling traffic that affects their bandwidth due to users' mobility pattern change.

To deliver service to large numbers of LTE subscribers with LTE User Equipment (UE) units, LTE networks need to know the exact location of the serving cell of each UE unit. In this circumstance, LTE networks must maintain track of the location of all UE units, and due to the high mobility of UEs, locating each UE's specific location (i.e., each UE's serving cell) within the network coverage area is difficult. The signaling load associated with tracking UE's is extremely high and amounts to 34 % of the total signaling load on the mobility management entity (MME) of the LTE network [2]. One of the primary functions of LTE networks is mobility management (MM). The network element that controls and manages an individual's UE's mobility within the LTE network is known as the MME and is tasked to handle all mobility-related messages between the UEs and the serving network [3].

The number of cells allocated to each TA is a compromise between the number of area updates and the volume of paging messages, or between uplink and downlink signaling burden, from the standpoint of planning and optimization [7]. Some of the factors that should be considered for TAU optimization are UE spatial density and mobility, whereas paging is dependent on traffic load characteristics of each cell [9]. In the current configuration of the LTE network of Ethiotelecom, the MME stores the TA in which the UE is registered in the TA update (TAU) and uses it to page or track a UE in the network area, which means that the location management scheme in place, in Ethiotelecom, is the standard scheme i.e. MME transmits a paging message in all cells of the UE's registered TA to place the call to the UE, and when the user changes a TA the user's location is updated. The current TA configuration, in the LTE network of Addis Ababa, is designed and optimized for location management in the planning phase which implies that it will no longer perform well as UE distribution and mobility patterns change over time. As a result, a TA redesign and reconfiguration may be necessary to reduce total signaling overhead. According to Jessica and Mario [6], the signaling overhead arising from a static configuration can be reduced significantly by using Tracking Area List (TAL) for location management. Nevertheless, such a method has not been investigated and applied for the Addis Ababa LTE network context. For large networks, like Addis Ababa's LTE network the design of TALs requires an underlying collection of TAs that have previously been planned. This thesis intends to carry on a TA design and optimization approach using distance-based clustering (K-means) and density-based clustering algorithms, Hierarchical Density-Based Spatial Clustering of Application with Noises (HDBSCAN), to produce optimal number of TAs that can used for efficient TAL implementation [10].

II. Related Works

Many technologies have been discussed about location management and many researchers have investigated the induced signaling overhead as a result of users' mobility. TAU and paging, in particular, have been identified as two key components of the location management method. Several investigations, like in [11], [12], and [13] potential alternative strategies for TAU signaling overhead minimization derived from traditional techniques such as timing, speed, and mobility for triggering the TAU procedure, thereby reducing signaling overhead. The TAL concept has intrigued scholars from both academia and the industry due to its preferred adaptability. TAL configuration study can be divided into two main categories: static and dynamic configuration. Several researchers looked at the TAL in various methods, with some using traditional optimization approaches connected to the notion of TAL in previous works [5], [13], [14].

In Razavi et al. [14], the researchers planned and evaluated by comparing a new TAL design approach to the conventional TA concept. The researchers also confirmed the significant dissimilarities in cell distribution between TA and TAL in a static mode. The same researchers suggested a dynamic TAL design in associated work [13], in which the cells are dynamically allocated to the system's existing TALs configurations. The rule of thumb approach was developed to adapt to the design of TAL in order to minimize the signaling overhead caused by TAU and paging, and it was used to allocate cells to tracking area lists. The study's findings show that the proposed dynamic approaches outperform the standard TA and static TAL in terms of performance results. Furthermore, the authors proposed a study that would use a linear programming model to minimize signaling load [5]. The TAU and paging signaling overhead regarding TAL creation were explored in two studies by Aqeeli et al. [8], [11], to link the MME with TAL. In both experiments, a centralized and a distributed MME technique were proposed. Each TAL construction is linked to one MME in the centralized system, but each TA that characterizes each cell is linked to one MME in the distributed scheme. Both types of research examine the implications of having various MME distributions throughout the system and the overall signaling overhead consequence. The first work [11], in contrast, provided a static technique for allocating cells to tracking area lists in a fixed period. When compared to the distributed MME technique, the centralized MME technique provides a more streamlined signaling overhead. The second study [8] presented an intelligent dynamic procedure that restructures cells among the TALs in a preset time based on a smart selection scheme. The smart selection strategy aids in the inclusion of cells in the same tracking area list that are more likely to have big connections between them. The research discovered a substantial improvement in total signaling overhead.

Jessica and Mario [9], discuss that the design of TALs requires an underlying set of TAs that has already been planned, though some TAL optimization algorithms assume each TA is just made up of one cell. However, in networks using current standards, assuming a TA to be only one cell may not be ideal. As a result, a K-means clustering algorithm is utilized to generate TAs to find the best Voronoi distribution of TAs that will be employed by the TAL optimization [9].

III. Methodology

The essential contextual information about TA design and optimization in general and distance-based clustering as well as density-based clustering algorithms are covered through literature reviews from journals and books. Call detail records data is collected from Ethiotelecom Customer Relation Management (CRM), infrastructure data (LTE network design data) was collected from the wireless Engineering division, and for map matching, Openstreet map data is downloaded and used for filling the location gaps observed in the CDR data. A laptop computer with Python environment with libraries: Geopandas, OSMnx, NetworkX, Shapely, scikit-mobility and other necessary tools are used. The methodologies followed to accomplish the general objective as well as the specific objectives of this thesis are depicted in Figure 1.2 below.

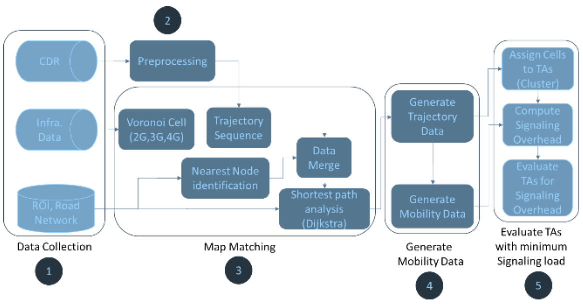


Figure 1.1 Analysis Procedure (Methodology)

IV. Scenario

The LTE network studied has 650 eNodeBs and a total of 930,774 unique LTE users are identified from the CDR data out of which 1% is randomly sampled. The sampled, 9307 users have 574,610 records in the CDR data collected over a 20-day duration. This data is transformed using map-matching techniques (shortest path analysis) on the drivable road network of Addis Ababa, as shown in Figure 1.3 below, to produce UE traces.

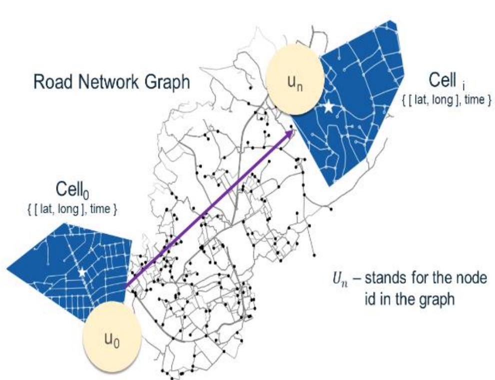


Figure 1.2 Voronoi Cell centroid mapping to the road network intersections

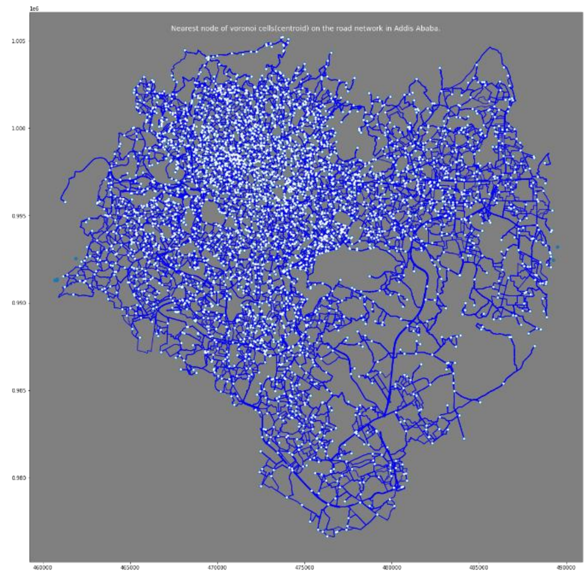


Figure 1.3 Final Road Network Graph for Shortest Path Analysis

The final reconstructed UE traces sample is shown in Table 1.1 below. This data is used to estimate the distribution and the mobility pattern of UEs.

Table 1. 1 UE traces for the LTE users in Addis Ababa (Reconstructed sample)

	CALLER ID	START TIME	U	LOCATION ID	LONGITUDE	LATITUDE
5482540	fffbb3c5c2e8	5/1/2021 19:25	262627495	1133	38.76247568	8.988843786
5482541	fffbb3c5c2e8	5/1/2021 19:37	6307707827	1536	38.75012871	8.992404156
5482542	fffbb3c5c2e8	5/1/2021 19:41	299009340	112201	38.75883528	8.991407367
5482543	fffbb3c5c2e8	5/1/2021 19:43	560335266	1299	38.75072041	8.998068166
5482544	fffbb3c5c2e8	5/1/2021 19:49	48613467	1356	38.74941968	9.008796636

The UEs distribution computed from the reconstructed UE traces is shown in figure 1.4 below

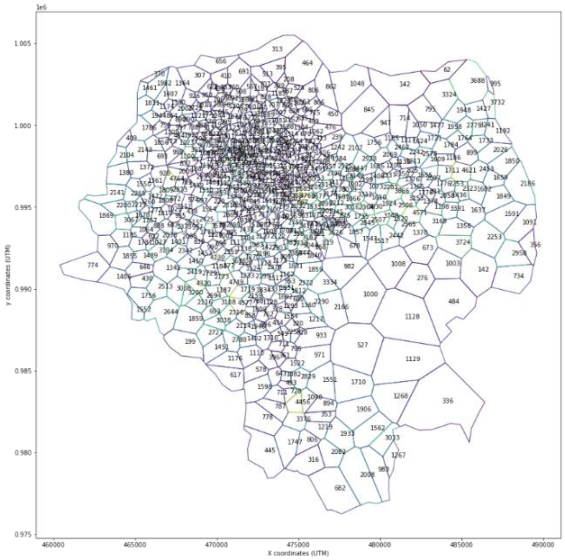


Figure 1.4 UE distribution for each Voronoi tessellation (eNodeB level)

V. Objective Function and Algorithms

The minimization objective function and analysis approach is shown below:

$$\underset{\mathbf{T}}{minimize} \quad S_{cost}(\mathbf{T}) = \alpha \times TAU_{cost}(\mathbf{T}) + \beta \times P_{cost}(\mathbf{T})$$

$$subject\ to \qquad N_{TA} \le N_{cell},$$

$$1 \le \sum_{i=1}^{N_{cell}} t_{ik} \le N_{cell}, \forall k,$$

$$\sum_{k=1}^{N_{TA}} t_{ik} = 1, \forall i$$

Where,

The tracking area update overhead defined as:

$$TAU_{cost}(\mathbf{T}) = \sum_{i=1}^{N_{cell}} \sum_{j=1}^{N_{cell}} (\mathbf{1} - X_{ij}) \times F_{ij} \times t_c$$

The paging overhead defined as:

$$P_{cost}(\mathbf{T}) = \sum_{k=1}^{N_{TA}} ((\partial \times \boldsymbol{N_{UE}}) \, . \, \mathbf{X_T} \, . \sum_{i=1}^{N_{cell}} t_{ik}) \times p_c$$

VI. Algorithm 1a

The algorithm embedding k-means clustering for assigning cells to TAs.

Algorithm 1a: Distance based clustering Algorithm (K-means) for allocation of cells to TAs

Input:

$\{\beta, \alpha, \partial, p_c, t_c, N_{cell}\}$

$x \in \mathbb{R}^{N_{cell} \times 2}$: coordinates of all cells

$N_{UE} \in \mathbb{N}^{N_{cell} \times 1}$: number of UEs in each cell

$\mathbf{F} \in \mathrm{N}^{N_{cell} \times N_{cell}}$. the number of UEs moving from cell i to cell j during the evaluation time

Output:

$\{\mathbf{T}^*, S^*, N_{TA}^*\}$

for n=1 : N_{cell} do

// Smartly guess n number of centroids

$\mu \leftarrow kmeans(x, n)$

//Get cluster labels

$\{T'\} \leftarrow kmeans(x, n, \mu)$

//Get unique number of labels which actually is n

$N_{TA}^* = count(unique\ cluster\ labels\ in\ \boldsymbol{T}')$

$X_T \leftarrow Evaluate\ same\ TA\ membership\ (\boldsymbol{T}')$

$TAU_{cost} \leftarrow Compute\ TAU\ cost((\mathbf{1} - X_T), \mathbf{F}, t_c)$

$P_{cost} \leftarrow Compute\ Paging\ cost(X_T, \partial, N_{UE}, p_c)$

$S_{cost} \leftarrow Compute\ total\ signaling\ cost(TAU_{cost}, P_{cost}, \beta, \alpha,)$

if n=1 **then**

//Compute the baseline signaling cost

$\{\mathbf{T^b}, S^b, N_{TA}^b\}$

$\{\mathbf{T}', S_{cost}, n\}$

//compare and select the minimum of the two

//compare and select the minimum of the two

$\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \min(\{\mathbf{T^b}, S^b, N_{TA}^b\}, \{\mathbf{T}', S_{cost}, n\})$

else if $S_{cost} < S^*$ **then**

$\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \{\mathbf{T}', S_{cost}, n\}$

end

VII. Algorithm 1b

The algorithm embedding HDBSCAN clustering for assigning cells to TAs.

Algorithm 1b: HDBSCAN clustering for allocation of cells to TAs

Input:

$\{\beta, \alpha, \partial, p_c, t_c\}$

$x \in \mathbb{R}^{N_{cell} \times 2}$: coordinates of all cells

$R_C \in \mathbb{R}^{N_{nodes} \times 2}$: the set of routes with coordinates of the cells for all UEs

$N_{UE} \in \mathbb{N}^{N_{cell} \times 1}$: number of UEs in each cell

$\mathbf{F} \in \mathbb{N}^{N_{cell} \times N_{cell}}$: the number of UEs moving from cell i to cell j

$k - the\ minimum\ size\ grouping\ to\ consider\ as\ a\ cluster - density$

$K \in \mathbb{N} - list\ of\ k - variable$

$s = 1: - minimum\ number\ of\ samples$

$\varepsilon = 700\ m\ -\ the\ distances\ between\ cells\ coordinate\ to\ stop\ separating\ clusters\ that\ are\ less\ than\ \varepsilon$

Output:

$\{\mathbf{T}^*, S^*, N_{TA}^*\}$

for k in K do
  //Initialize iteration number
  $i = 0$
  //Get cluster labels
  $\{T'\} \leftarrow HDBSCAN(\boldsymbol{R_r}, k, s, \varepsilon)$
  //Get the unique number of labels
  $N_{TA}^* = count(unique\ cluster\ labels\ in\ \boldsymbol{T'})$
  $\boldsymbol{X_T} \leftarrow Evaluate\ same\ TA\ membership\ (\boldsymbol{T'})$
  $TAU_{cost} \leftarrow Compute\ TAU\ cost((\mathbf{1} - \boldsymbol{X_T}), \mathbf{F},\ t_c)$
  $P_{cost} \leftarrow Compute\ Paging\ cost(\boldsymbol{X_T}, \partial, \boldsymbol{N_{UE}}, p_c)$
  $S_{cost} \leftarrow Compute\ total\ signaling\ cost(TAU_{cost}, P_{cost}, \beta, \alpha)$

  if n=1 then
    //Compute the baseline signaling cost
    $\{\mathbf{T}^b, S^b, N_{TA}^b\}$
    $\{\mathbf{T}', S_{cost}, n\}$ //compute the new signaling cost
    //compare and select the minimum of the two
    $\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \min(\{\mathbf{T}^b, S^b, N_{TA}^b\}, \{\mathbf{T}', S_{cost}, n\})$
  else if $S_{cost} < S^*$ then
    $\{\mathbf{T}^*, S^*, N_{TA}^*\} \leftarrow \{\mathbf{T}', S_{cost}, n\}$
  $i = i + 1$ // increment the iteration number by one
end

VIII. Results and Discussions

The parameters and values shown in Table 1.2 have been used for the signaling overhead analysis.

Table 1. 2. Parameters and Values

Parameters	Values	Remarks
Number of Cells (N_{cell})	650	Number eNodeBs
Baseline number of TAs (N_{TA}^b)	8	Number of TAs for the current configuration
Paging rate ∂	20%	Due to the data usage and push notification [6]
Paging factor β	1	
Tracking update factor $\boldsymbol{\alpha}$	10	TAU to paging cost (10:1), due to UE battery utilization and battery life [6]
Paging cost , p_c	33 messages	33 messages for one paging attempts [2]
Tracking update cost t_c	40 messages	40 messages for one tracking area update [2]

The signaling overhead performance evaluation of all the approaches is presented in Table 1.3 below and the following discussion is based on that.

Table 1. 3. Parameters and Values

Algorithm	Density	Number of TAs	Tracking area update cost $TAU_{cost}(\boldsymbol{T})$	(%)	Paging cost $P_{cost}(\boldsymbol{T})$	(%)	Total signaling cost $S_{cost}(\boldsymbol{T})$	Improvement (%) $\frac{S_{cost}(\boldsymbol{T}^b) - S_{cost}(\boldsymbol{T}^*)}{S_{cost}(\boldsymbol{T}^b)}$ (%)
Baseline		8	1.30E+08		4.78E+08		6.08E+08	
One Cluster		1	0.00E+00	100%	3.96E+09	-728%	3.96E+09	-551.32%
K-means		74	2.54E+08	-95%	5.31E+07	89%	3.07E+08	49.51%
HDBSCAN	400	237	2.38E+08	-83%	7.79E+06	98%	2.46E+08	59.54%
Single Cell TAs		650	3.71E+08	-185%	6.10E+06	99%	3.77E+08	37.99%

The overall signaling overhead of the baseline TAs (8) configuration is found to be $S_{cost}(\mathbf{T}) = 6.08\text{E} + 08$ which has a TAU component of 1.30E+08 messages and a paging component of 4.78E+08. The results for TAs created with K-means clustering approach shows that the worst case in terms of the signaling cost is when the allocation of 650 cells into one TA is done, i.e., N_{TA} = 1, producing 551.32 % more overall signaling overhead as compared to the baseline overall signaling overhead, as the contribution of the P_{cost} is 100 %, even though the TAU component is effectively zero. Creating TAs having a single cell as a member, i.e., N_{TA} = 650, has 99 % improvement for paging overhead but produces 185 % more tracking area update overhead, which results in a 37.99 % overall signaling overhead improvement as compared to the baseline. The grouping of the cells into 74 TAs (N_{TA} = 74) shows that the tracking area update overhead is 95 % more than the baseline TAU overhead while the paging overhead is 89 % less than the corresponding baseline paging signaling overhead, though the total overall minimum signaling cost $S_{cost}(\mathbf{T}) = 3.07\text{E}+08$ has 49.45 % improvement as compared to the baseline overall signaling overhead.

Similarly, the results for TAs created with HDBSCAN clustering approach (N_{TA} = 237) shows that the TAU overhead is 83 % more than that of the baseline TAU overhead while the paging overhead is 98 % less than that of the baseline paging overhead, though the total overall minimum signaling cost $S_{cost}(\mathbf{T}) = 2.46\text{E}+08$ has 59.54 % improvement as compared to the baseline overall signaling overhead. Therefore, TAs created with HDBSCAN clustering approach has a better performance (59.54 %) in terms of the total signaling overhead against the baseline TAs, due to the fact that HDBSCAN considers the density of the LTE users while grouping cells into TAs. Figure 1.5 shows the total signaling overhead trend vs the number of TAs (aligned with the same N_{TA}): for the baseline, for TAs created with K-means clustering approach and for TAs created with HDBSCAN clustering approach.

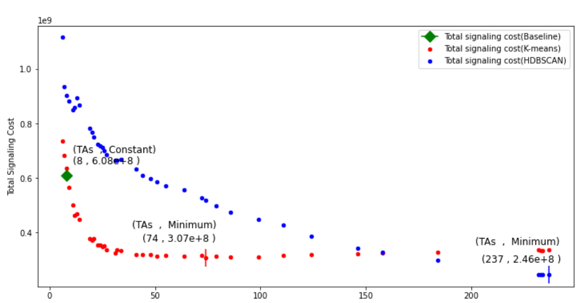


Figure 1.5 Total signaling overhead trend (aligned with the same N_{TA})

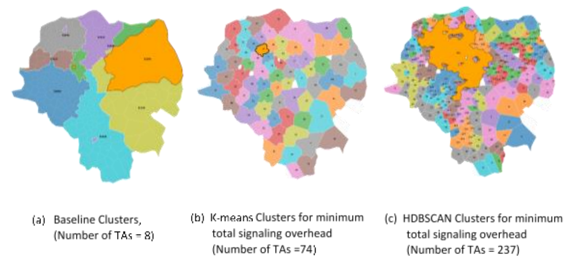


Figure 1.6 Cell distribution (for the minimum total overall signaling overhead)

Figure 1.6 shows the cell distribution for the minimum total signaling overhead for the baseline, for TAs created with K-means clustering approach and for TAs created with HDBSCAN clustering approach.

IX. Conclusion

The tracking area design and optimization approach for location management signaling overhead in LTE networks is confirmed to adjust to changes and trends of UE location and mobility patterns, though the sample data used for analysis may have some bias. For real networks having a large number of cells, creating TAs with single-cell member-only or assigning all cells to a single TA, or keeping the TAs deployed during initial stages may not be optimum for signaling overhead. To tackle the problem of allocating cells to TAs in such a way that the overall signaling overhead is minimized a heuristic algorithm, independently and separately embedded with two clustering techniques, has been implemented. In the first heuristic algorithm, the K-means clustering technique is embedded within it and produces relatively good results, as compared to the baseline. In the second heuristic algorithm, the HDBSCAN clustering technique is embedded within it and results are better than the baseline as well as the K-means clustering technique, as HDBSCAN takes the density and mobility pattern of the users into account to assign cells to TAs. The case study on a realistic network using the standard location management schemes shows that the heuristic algorithm embedded with the HDBSCAN clustering technique is able to improve the signaling overhead significantly.

X. Recommendations

As the sampling technique employed is random (i.e. user profiling is not used), the performance results of the analysis may have some bias. Taking the bias in mind, the following recommendations are made:

- The case study network location management signaling overhead can be improved significantly with the use and implementation of the TA design and optimization approach presented in this thesis.
- For efficient dynamic TAL implementation, the TA design and optimization approach presented in this thesis can be used as a foundation.
- The work in this thesis can be extended by considering algorithms that assign TA into TAL and implement TAL assignment to individual UEs in the LTE network through the incorporation of individual mobility patterns and incoming call connection characteristics.

XI. References

[1] Ericsson, "Ericsson Mobility Report," SE-164 80 Stockholm, Sweden, 2021.

[2] Alaa A. R. Alsaeedy and Edwin K. P. Chong, "A review of mobility management entity in LTE networks," *Int J Network Mgmt,* no. e2088, 2019.

[3] 3GPP TS 23.401 v14.4.0., *General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access (release 14),* 2017.

[4] Aqeeli E, Hashim HA, Haque A and Shami A, "Optimal location management in LTE networks using evolutionary techniques," *Int J Commun Syst,* vol. e4100, 2019.

[5] Jessica Moysen and Mario Garc´ıa-Lozano, "Learning-based Tracking Area List Management in 4G and 5G Networks," *in IEEE Transactions on*

Mobile Computing, vol. 19, no. 8, pp. 1862-1878, 1 Aug 2020.

[6] Github, "Understanding HDBSCAN and Density-Based Clustering," 17 Jan 2021. [Online]. Available: https://pberba.github.io/stats/2020/01/17/hdbscan/. [Accessed 15 Dec 2021].

[7] Emad Aqeeli, Abdallah Moubayed and Abdallah Shami, "Towards Intelligent LTE Mobility Management through MME Pooling," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015.

[8] Kuo Hsing Chiang and Nirmala Shenoy, "A 2D Random Walk Mobility Model for Location Management Studies in Wireless Networks," *IEEE Trans Veh Technol,* vol. 53, no. 2, p. 413–424, 2004.

[9] Modarres Razavi S and Yuan D, "Mitigating mobility signaling congestion in LTE by overlapping tracking area lists," in *Proceedings of the 14th acm international conference on modeling, analysis and simulation of wireless and mobile systems*, New York, 2011.

[10] Razavi SM, Yuan D, Gunnarsson F and Moe J, *Exploiting tracking area list for improving signaling overhead in LTE,* 2010, pp. 1-5.

[11] Sara Modarres Razavi, Di Yuan, Fredrik Gunnarsson and Johan Moe, "Dynamic Tracking Area List configuration and performance evaluation in LTE," *2010 IEEE Globecom Workshops,* pp. 49-53, 2010.

[12] E. Aqeeli, A. Moubayed and A. Shami, "Dynamic SON-Enabled Location Management in LTE Networks," *in IEEE Transactions on Mobile Computing,* vol. 17, no. 7, pp. 1511-1523, 1 Jul 2018.

[13] Sara Modarres Razavi, *Tracking Area Planning in Cellular Networks - Optimization and Performance Evaluation,* Sweden: Link¨ oping Studies in Science and Technology, 2011.

[14] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu and Wenye Wang, "Mobility management in next-generation wireless systems," *in Proceedings of the IEEE,* vol. 87, no. 8, pp. 1347-1384, Aug 1999.

[15] Ethiotelecom, *Addis Ababa's LTE network plan,* Addis Ababa: Ethiotelecom, 2021.

[16] I. Demirkol, C. Ersoy, M.U. Caglayan and H. Delic, "Location area planning in cellular networks using simulated annealing," *IEEE Conference on Computer Communications,* vol. INFOCOM, no. 01, pp. 13-20, 2001.

[17] Daqing Gu and Stephen S. Rappaport, "Mobile user registration in cellular systems with overlapping location areas," *1999 IEEE 49th Vehicular Technology Conference (Cat. No.99CH36363),* vol. 1, pp. 802-806, 1999.

[18] T. Winter, "Mobility management and network design for UMTS," in *Proc. of IEEE International Symposium on Personal, Indoor and*, 2004.

[19] GeoPandas, "The documentation of GeoPandas," The documentation of GeoPandas, [Online]. Available: https://geopandas.org/en/stable/. [Accessed 1 November 2021].

[20] Humdata (HDX), "Ethiopia - Subnational Administrative Boundaries," November 2021. [Online]. Available:

https://data.humdata.org/dataset/ethiopia-cod-ab. [Accessed 01 November 2021].

[21] OpenStreetMap, "Open Street Map," OpenStreetMap, 1 November 2021. [Online]. Available: https://www.openstreetmap.org/#map=13/9.0174/38.8372. [Accessed 1 November 2021].

[22] OSMnX, "User reference for the OSMnx package.," OSMnX, [Online]. Available: https://osmnx.readthedocs.io/en/stable/osmnx.html. [Accessed 1 November 2021].

[23] towardsdatascience, "Understanding K-means Clustering in Machine Learning," 13 Sep 2018. [Online]. Available: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1. [Accessed 15 Dec 2021].

[24] J. O. Abolade, O. A. Fakolujo and A. Orimogunje, "Handover in Mobile Wireless Communication Network - A Review," *Advanced Engineering, management and Science,* vol. 3, no. 9, p. 934–940, 2017.

[25] Dahlman E, Parkvall S and Sköld J, 4G: LTE/LTE-advanced for Mobile Broadband, Second ed., Chennai: Elsevier/Academic Press, 2011.

[26] J. Cowling, *Dynamic location management in heterogeneous cellular networks,* University of Sydney, 2004.

[27] NetworkX, "Networkx Algorithms Shortest Paths Weighted Bidirectional Dijkstra," NetworkX, [Online]. Available: https://networkx.org/documentation/networkx-1.8/reference/generated/networkx.algorithms.shortest_paths.weighted.bidirectional_dijkstra.html. [Accessed 1 November 2021].

[28] Chaoming Song, Tal Koren, Pu Wang and Albert-László Barabási, "Modelling the scaling properties of human mobility," *Nature physics,* vol. 6, 13 September 2010.

[29] Pappalardo, L., Simini and F, "Data-driven generation of spatio-temporal routines in human mobility," *Data Min Knowl,* vol. Disc 32, p. 787–829, 2018.

[30] Scikit-mobility, "Scikit-mobility," Scikit-mobility, [Online]. Available: https://scikit-mobility.github.io/scikit-mobility/. [Accessed 1 November 2021].

Appendix B

UE Distribution for the sample Space, from the CDR (1 of 3)

Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs
111001	9.61	111052	13.79	111115	18.23	111168	11.39	111221	36.00
111002	10.96	111053	10.39	111117	8.93	111169	12.85	111222	24.40
111003	4.96	111054	17.68	111118	9.05	111170	20.21	111223	18.25
111004	12.90	111055	13.05	111120	18.20	111171	6.87	111224	14.07
111005	8.22	111056	14.08	111121	8.60	111172	25.75	111225	47.78
111006	13.76	111057	20.67	111122	14.58	111173	11.34	111226	14.09
111007	11.68	111063	8.69	111123	15.26	111174	47.94	111227	10.33
111008	23.99	111064	37.00	111124	9.55	111175	25.31	111228	23.30
111009	29.21	111065	28.07	111125	15.26	111176	6.76	111230	21.10
111010	50.76	111066	7.65	111415	7.60	111177	22.12	111231	8.41
111011	3.60	111068	7.44	111127	14.36	111178	11.45	111267	33.97
111012	0.62	111069	10.05	111128	28.83	111179	7.03	111269	4.36
111013	9.68	111070	13.01	111129	15.25	111180	21.27	111271	18.70
111015	13.08	111071	18.42	111130	6.26	111181	15.19	111272	8.24
111016	18.08	111073	15.31	111131	10.98	111182	9.42	111273	1.49
111017	10.73	111074	16.98	111132	3.98	111183	41.56	111274	13.59
111018	31.96	111075	6.14	111133	17.17	111184	10.71	111275	5.66
111020	25.98	111076	6.73	111134	27.72	111185	11.26	111276	16.32
111021	34.38	111077	10.52	111135	8.80	111187	16.68	111277	23.28
111022	18.99	111078	10.75	111136	20.17	111188	14.46	111278	12.15
111023	20.00	111080	15.28	111137	6.71	111189	10.28	111279	32.73
111024	16.60	111081	6.80	111139	8.56	111190	17.82	111388	20.44
111025	4.86	111082	23.51	111140	42.28	111191	4.40	111389	7.63
111026	9.22	111083	14.56	111141	10.06	111192	18.44	111390	40.27
111027	9.02	111085	23.13	111142	4.86	111193	11.80	111393	15.19
111028	24.81	111087	16.66	111143	15.00	111195	10.12	111394	19.84
111029	11.65	111088	33.32	111144	20.54	111196	7.15	111395	14.82
111030	22.15	111089	25.23	111145	7.05	111197	8.93	111396	3.39
111031	10.69	111090	13.27	111146	19.20	111198	7.64	111397	31.14
111032	11.03	111091	12.18	111147	4.66	111199	6.61	111399	18.51
111033	7.24	111092	13.73	111148	5.17	111200	10.42	111410	21.56
111034	12.26	111094	20.81	111149	4.91	111201	11.96	111403	6.53
111035	8.95	111095	17.46	111150	5.39	112265	4.02	111404	5.04
111036	10.87	111096	25.94	111151	6.46	111203	11.24	111405	47.90
111037	6.48	111098	23.64	111152	27.46	111204	11.69	111406	6.25
111038	12.14	111100	25.21	111153	7.43	111205	5.81	111411	15.77
111040	12.47	111101	13.86	111154	3.55	111206	19.75	111414	10.71
111041	13.64	111104	6.17	111155	6.82	111210	10.05	111415	8.51
111042	26.99	111106	19.35	111156	17.30	111213	14.68	111417	10.11
111043	13.11	111107	22.97	111157	13.18	111214	5.77	111418	6.50
111044	9.16	111109	11.34	111158	12.28	111215	20.98	111419	14.93
111045	10.01	111110	14.80	111159	4.90	111216	11.24	111420	7.52
111046	6.93	111111	15.64	111160	10.03	111217	21.35	111421	22.26
111047	17.94	111112	16.88	111164	18.68	112077	6.42	111422	9.59
111049	16.21	111113	25.41	111166	46.00	111219	9.96	111423	10.63
111051	9.56	111114	12.11	111167	3.00	111220	8.44	111425	12.50

UE Distribution for the sample, Space from the CDR (2 of 3)

Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs
111427	28.91	112093	20.67	111062	11.81	111546	20.51	111633	9.04
111428	7.91	112114	4.99	111067	8.10	111547	5.38	111635	2.40
111429	14.11	112095	7.60	111072	12.25	111549	27.45	111637	22.68
111430	32.13	112096	5.49	111079	6.95	111550	10.03	111638	19.94
111431	11.23	112097	8.93	111084	11.24	111551	11.07	111642	12.02
111432	12.67	112098	12.45	111086	17.22	111553	15.23	111643	8.15
111434	14.26	112099	5.85	111093	17.32	111555	46.06	111644	10.79
111558	22.13	112101	9.13	111097	11.93	111557	9.88	111646	14.17
111560	12.28	112102	8.99	111099	10.80	111559	18.46	111647	16.20
111565	19.46	112103	10.11	111102	22.67	111561	11.15	111648	19.56
111618	9.47	112104	14.43	111103	46.50	111562	12.31	111649	34.12
111024	13.57	112105	18.44	111105	13.44	111563	6.25	111650	8.30
111631	2.76	112108	8.92	111108	8.74	111564	5.84	111651	39.37
111634	36.49	112109	7.45	111119	17.75	111567	6.46	111652	11.90
111639	30.51	112115	11.97	111138	10.24	111569	10.57	111654	5.54
111645	15.70	112111	6.10	111161	7.19	111570	4.99	111655	5.49
111653	15.17	112112	19.90	111162	20.71	111571	18.19	111657	23.87
111656	8.62	111822	16.13	111163	14.15	111574	28.05	111658	28.46
111687	11.30	111886	23.97	111165	24.68	111576	11.94	111659	24.77
111772	7.68	112094	17.68	111186	14.98	111577	43.47	111661	8.00
112110	5.72	112116	7.03	111194	4.12	111578	12.40	111662	15.93
111866	23.24	112117	13.52	111207	3.63	111579	21.87	111663	13.93
111883	27.50	112118	7.94	111211	30.26	111580	20.15	111664	11.94
111884	13.39	112119	23.16	111212	27.10	111582	17.87	111665	10.13
111885	36.57	112120	12.56	111229	17.91	111583	12.05	111666	11.34
112113	30.35	112121	21.35	111244	3.71	111584	4.32	111667	6.91
112070	22.55	112122	5.56	111268	9.77	111585	9.60	111668	9.82
112071	7.70	112123	9.86	111270	13.68	111586	14.61	111669	21.20
112073	7.57	112124	5.71	111273	6.88	111590	17.95	111670	5.02
112074	12.94	112139	8.12	111300	14.70	111593	5.65	111673	15.72
112076	12.26	112140	7.78	111387	33.55	111598	7.92	111674	30.52
112077	25.04	112239	12.77	111398	16.34	111599	5.27	111676	6.97
112078	6.68	112264	2.15	111400	31.88	111600	4.78	111678	6.38
112079	10.31	112265	0.42	111402	9.29	111603	10.08	111679	2.96
112080	13.53	112266	1.12	111407	22.00	111604	8.10	111680	9.33
112081	13.65	115001	2.74	111408	20.95	111609	24.57	111681	18.81
112082	11.46	111012	6.92	111409	23.37	111610	14.86	111682	7.25
112083	8.69	111014	18.27	111412	9.28	111613	16.18	111683	7.11
112084	33.46	111019	4.49	111413	37.56	111617	9.11	111684	5.11
112085	11.65	111039	8.33	111416	4.26	111623	5.74	111685	5.99
112086	24.24	111048	13.72	111424	19.77	111624	17.87	111686	12.51
112087	32.17	111050	2.96	111426	17.97	111626	44.84	111689	6.25
112088	9.14	111058	7.34	111433	11.92	111627	12.25	111690	6.50
112089	18.43	111059	19.35	111435	27.96	111628	12.13	111696	1.88
112090	14.50	111060	5.70	111436	6.85	111629	16.65	111697	7.94
112092	9.05	111061	4.01	111439	3.15	111632	31.60	111698	10.03

UE Distribution for the sample Space, from the CDR (3 of 3)

Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs	Site ID	Number of UEs
111699	11.18	111843	7.10	112137	8.99	112193	30.57	117107	31.27
111700	17.68	111844	7.64	112138	7.24	112194	8.00	117108	25.29
111701	4.71	111845	15.81	112141	19.17	112195	19.60	117111	16.23
111702	13.81	111846	6.95	112142	20.20	112197	0.62	117112	22.30
111708	12.68	111848	8.23	112143	8.67	112198	3.55	117116	22.56
111710	13.69	111849	5.16	112144	18.60	112199	20.38	117117	11.45
111719	25.81	111851	13.12	112145	17.43	112200	16.69	117120	21.17
111720	13.84	111852	3.97	112146	16.47	112201	9.97	117134	9.76
111721	9.26	111853	4.52	112147	21.79	112202	10.17	117137	32.20
111722	15.41	111855	4.15	112148	19.45	112205	10.55	117138	14.59
111726	14.45	111856	6.08	112149	18.61	112208	16.93	117139	13.18
111729	10.88	111858	29.46	112150	7.38	112209	17.58	117140	12.56
111732	3.62	111859	17.67	112151	5.29	112210	6.59	117141	6.93
111733	7.15	111861	14.15	112152	10.98	112211	21.73	117144	21.54
111735	6.78	111862	16.97	112155	7.18	112212	6.21		
nan	5.10	111864	3.09	112156	18.67	112213	3.17		
111738	4.47	111865	3.55	112157	2.00	112214	2.78		
111742	18.67	111867	6.85	112158	10.13	112215	5.81		
111743	5.16	111874	9.32	112159	17.67	112220	19.79		
111744	8.95	111876	6.27	112160	4.67	112222	29.77		
111745	16.72	111881	3.30	112161	9.53	112224	7.85		
111746	10.03	111882	7.67	112163	18.80	112225	13.80		
111747	8.94	111890	12.76	112164	7.91	112227	1.42		
111749	18.51	111896	1.43	112165	15.60	112228	12.75		
111750	11.91	111999	9.88	112166	11.83	112229	18.70		
111751	24.33	112037	37.11	112167	27.44	112230	37.47		
111752	23.57	112048	22.10	112168	9.57	112234	7.78		
111758	23.79	112059	26.60	112169	6.97	112235	13.88		
111759	6.73	112062	15.33	112170	4.11	112236	16.01		
111762	2.09	112069	8.57	112171	6.77	112244	17.21		
111763	8.71	112072	5.20	112173	13.64	nan	0.46		
111765	11.16	112075	9.02	112175	10.06	112246	11.99		
111767	5.52	112091	12.19	112176	4.87	112248	23.35		
111773	11.36	112100	15.41	112178	8.50	112250	10.01		
111806	32.61	112106	10.85	112179	15.59	112251	3.58		
111809	33.45	112126	19.70	112180	16.07	112252	28.63		
111819	13.50	112127	18.42	112181	3.38	112257	22.67		
111821	18.20	112128	13.38	112182	30.86	112258	9.88		
111832	7.73	112129	19.58	112183	12.20	112263	10.09		
111833	8.63	112130	37.73	112184	21.29	113000	5.46		
111835	3.22	112131	38.54	112186	14.14	113001	7.03		
111836	6.59	112132	38.90	112187	22.79	113002	6.76		
111837	7.77	112133	13.78	112188	23.05	114001	7.83		
111838	7.55	112134	14.35	112190	11.35	117101	6.83		
111841	6.34	112135	15.52	112191	9.39	117102	10.06		
111842	12.26	112136	11.92	112192	2.21	117106	15.62		