This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

A New Methodology for User Equipment Trajectory Prediction in Cellular Networks

J. Sánchez-González, O. Sallent, J. Pérez-Romero

***Abstract*— User mobility prediction can be exploited in cellular networks for different purposes, such as enhancing the handover process, proactive resource allocation, proactive load balancing, etc., in order to improve the network performance. While many works aimed for the prediction of the next cell visited by the User Equipment (UE), the prediction of future UE locations has received less attention. In fact, only a few works deal with the prediction of the next UE location while other few works aim to predict the future direction of UEs arriving at a crossroad. This paper presents a methodology for the prediction of the trajectory followed by the UEs inside the cell. First, UE trajectory patterns are learnt by means of an off-line clustering of historical UE trajectories. Then, the obtained trajectory patterns are used for on-line prediction of the UE trajectory inside the cell, the prediction of the next cell that the UE will visit and an estimation of the time to reach this new cell. A dataset with UE trajectories moving around a large real-life cellular network has been considered.**

Keywords*— *Cellular networks, UE mobility, clustering, trajectory prediction, next-cell prediction.

I. Introduction

Fifth generation (5G) and beyond mobile networks are expected to support a great variety of applications and services with challenging requirements in terms of high data rate, high data volumes, reliable and low-latency communications, low energy consumption, high user density, high user mobility, etc. To deal with these challenges, the density of deployed cells will progressively increase. As a result, due to the smaller cell coverage and the user mobility, a User Equipment (UE) will be frequently changing their serving cells, resulting in heavy signalling overheads, risk of connection droppings and degraded Quality of Service (QoS). This situation poses great challenges to the network operators for the management of the radio resources among the different cells. In this context, the availability of mechanisms to determine the UE geographical location and to report these location measurements to the network [1] enable the exploitation of this information by using (Big) Data analytics technologies [2][3]. In this respect, the use of AI/ML (Artificial Intelligence/Machine Learning) methodologies, e.g. for the prediction of the UE mobility, can enhance the network performance by taking more proactive actions in certain functionalities such as handover, resource allocation, load balancing, etc.

In the last years, many works have focused on next cell prediction (i.e. prediction of the next cell visited by a UE) [4-16]. However, very few solutions have been proposed for the prediction of the trajectory that a UE will follow inside a specific cell. Thus, the main contribution of this paper is to propose a new methodology for UE trajectory prediction based on two steps. In the first step, an off-line trajectory modelling is done for the characterization of UE trajectory patterns in a cellular network. As part of this modelling, a set of collected UE trajectories are pre-processed following an interpolation process. Then, the interpolated trajectories are used as input to a clustering process to obtain a list of prototype UE trajectories inside each cell and to identify the most frequent and relevant UE trajectories inside the cells. Several metrics are defined to assess the quality of the clustering process. Then, in the second step, the obtained trajectory model is used for the on-line prediction of the trajectory that a UE will follow inside a cell. This prediction is done by comparing the current trajectory of a UE with the different prototype trajectories of the cell. Based on this, the likelihood of the prediction is estimated. The proposed methodology can also predict the next cell visited by the UE and estimate the time to reach this new cell. Several metrics are defined to assess the quality of the prediction process in terms of the achieved accuracy.

Besides the formulation of the novel methodology, the proposed approach considers practicality aspects in the design and is aligned with the 3GPP functional framework for AI-enabled Radio Access Network (RAN) intelligence of [17], which constitutes another contribution of the work. Furthermore, it is worth highlighting that the proposed approach has been tested in a realistic scenario with hundreds of cells using the dataset of [18] that includes real-life mobility trajectories of more than 50 thousand vehicles moving around inside the city of Cologne. The methodology has been extensively evaluated in heterogeneous situations (cells covering areas in the city centre, suburban residential areas,

Manuscript submitted October 19, 2022. Revised August 16, 2023, and December 18, 2023. Accepted March 25, 2024. This work is part of VERGE project, which has received funding from the Smart Networks and Services Joint Undertaking under the European Union's Horizon Europe research and innovation programme under Grant Agreement 101096034. This paper has also been partly funded by the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033 under ARTIST project (ref. PID2020-115104RB-I00). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The authors are with the Department of Signal Theory and Communications at the Universitat Politécnica de Catalunya (UPC), 08034, Barcelona (Spain), (e-mail: juan.sanchez-gonzalez@upc.edu, jose.oriol.sallent@upc.edu, jordi.perez-romero@upc.edu).

highways, etc.) with different cell shapes, cell sizes, cell density, different user mobility patterns, etc.

The rest of the paper is organised as follows. In Section II, the literature related to UE mobility prediction is discussed. Section III presents the general network architecture and the proposed solution for UE trajectory prediction aligned with the 3GPP specifications. Section IV details the proposed clustering and prediction methodology. Section V describes the considered realistic scenario, while Section VI presents the obtained results. Finally, the conclusions are summarised in Section VII.

II. Related work

Different research works in the literature exploit UE mobility predictions for different application areas in cellular networks, aiming to improve the performance at different levels of the protocol stack [4]. A large amount of papers focus on next cell prediction and an estimation of the remaining time at the current cell. These kind of predictions can be useful e.g. to assist handover decisions [5]-[11]. Some of these works make use of historical records of previously visited cells by the UEs and dwell time spent at each cell to determine the handover probability to the different neighbour cells e.g. [6]-[9]. Other approaches, such as [10], include multiple predictive factors (e.g. cell geometry, terrain characterization, description of roads and intersections, etc.) in order to improve the prediction accuracy. Other proposed methodologies include signal strength measurements and mobility of adjacent vehicles to optimise the handover process, e.g. [11][12]. Next cell prediction is also useful for the optimization of load balancing processes in cellular networks, like in [13] that predicts UE mobility and estimates future cell loads in a heterogeneous network to proactively balance the traffic among small cells and macrocells. Another applicability area is the optimization of mobile edge computing and caching [14]-[16]. As an example, [15] makes use of the sequence of Mobile Edge Computing (MEC) servers visited by the UE to design a mobility-aware caching approach that manages more efficiently the contents stored at each edge computing server relying on the prediction of the future MECs visited by the UE. Different mobility prediction techniques have been proposed in the literature in the context of next-cell prediction. These include Markov models e.g. [6][7][13]-[15], Bayesian Networks [10], Neural Networks [9][11][12], Support Vector Machines (SVM) [16] or clustering techniques [8] that collect the historical sequence of cells visited by each UE to identify UE mobility patterns and apply the obtained knowledge for next cell prediction.

A more reduced number of papers have dealt with the prediction of future UE geographical locations. Among them, papers [19-22] focus on how the future UE location prediction can be exploited in a wireless network and evaluate its benefits by means of simulations using theoretical mobility models. One application area is proactive resource allocation, i.e. to exploit knowledge of future user mobility and future wireless channel conditions to carry out a resource allocation planning beforehand to improve the network performance in terms of throughput, energy efficiency, etc. [19]-[21]. As shown in [21], some benefits can be obtained by opportunistically transmitting more data when the UE is located in regions where the channel conditions are favourable. As an example, the context information obtained by route prediction of a UE with a streaming service may anticipate a coverage hole in the trajectory of the UE so that more resources can be allocated to this UE before running into the coverage hole. Then, buffered data can maintain the streaming service experience while the UE is in the coverage hole. Similarly, for a streaming UE located at the cell edge or in a region with poor coverage, the minimum amount of data may be transmitted while more resources may be allocated when this UE moves to the cell centre or to regions with more favourable propagation conditions. On the other hand, knowing the specific future UE trajectory inside the cell can also be useful for a more accurate prediction of the UE dwell time at the current cell and the time to handover to the next cell. This may be useful for improving the handover or load balancing processes. UE location prediction can also be useful for mobility management in Ultra-Dense Networks (UDN), like [23] that proposed a proactive solution for the activation of gNodeBs (gNBs) to form virtual cells based on an estimation of the future user location.

Only a few works have been proposed for the prediction of future UE locations. A few papers, such as [23]-[25], focus on the prediction of the next UE location by means of SVM or Neural Networks. However, the prediction of just one sample corresponding to the next UE location may not be enough for proactive resource allocation, handover or load balancing purposes. In turn, other works, such as [26][27], aim to determine the probabilities that the UE takes different directions when arriving at a crossroad. These techniques proposed in [26][27] are able to determine the street the UE will take when arriving at a crossroad but not the specific UE locations at specific future instants of time. The aim of this work is to fill this gap by proposing and evaluating a prediction methodology to determine the trajectory followed by the UE (i.e. the specific UE locations at specific future instants of time) until the UE reaches the next cell. To the best of the authors' knowledge, no previous paper has focused on such characterization of UE trajectories inside the coverage region of a BS and the prediction of the trajectory that the UEs will follow inside the cell.

III. Network architecture

Fig. 1 presents how the proposed mobility prediction approach can be deployed in the 5G system architecture. It is based on an initial training step in which prototype trajectory patterns are learnt by means of an offline clustering of historical UE trajectories. Then, a second inference step uses the obtained prototype trajectories for on-line trajectory prediction. According to the functional framework for an AI-enabled RAN defined by 3GPP in [17], the AI/ML model training function that is in charge of the offline clustering process is deployed in the OAM (Operation and Management) system. In turn, the

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

inference function in charge of the online UE trajectory prediction is run in the NG-RAN (New Generation Radio Access Network), and more specifically in the Central Unit (CU) of the base station, i.e. the gNB in the 5G system. The gNB-CU hosts the upper layers of the radio interface protocol stack, including the Radio Resource Control (RRC) layer in charge of UE mobility control. A gNB-CU is connected to one or multiple gNB Distributed Units (gNB-DU), each one supporting one or more cells. The gNB-DU hosts the lower layers including the Medium Access Control (MAC) layer in charge of the MAC scheduling. The output of the AI/ML model inference can be used as an input for different functionalities identified in [17], such as:

- *Mobility optimization (MO):* UE mobility prediction is useful for selecting the handover target cell to assure service-continuity during UE mobility by minimising Radio Link Failures and call droppings (caused by e.g. a too-late or too early handover or a handover to a wrong cell) and avoid unnecessary handovers or ping-pong effects.
- *Load balancing (LB):* UE trajectory prediction is useful to evenly distribute the load/traffic among the different cells and Radio Access Technologies.
- *MAC Scheduling:* UE trajectory prediction inside the cell is useful to execute a proactive resource allocation and opportunistically transmit more data when the UE is located in regions where the channel conditions are favourable.

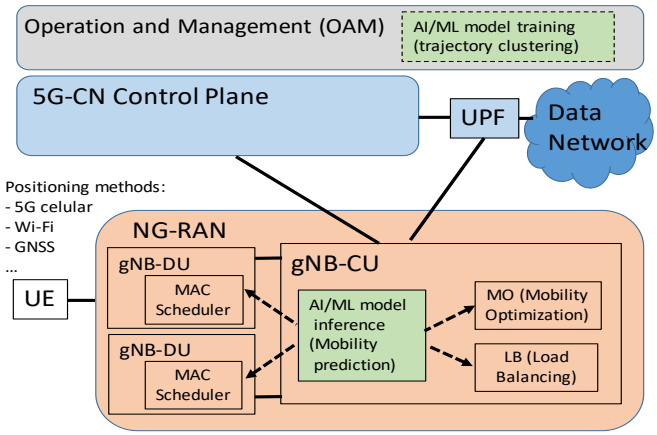


Fig. 1. 5G Network architecture.

Fig. 2 presents an example of flow diagram that illustrates how the proposed methodology can be used to predict a future UE handover from a cell covered by gNB 1 to a cell covered by gNB 2. The first part of the flow diagram shows the model training phase where the prototype trajectories in each gNB are learnt based on collected measurements of the UEs served by this gNB. For this purpose, whenever a UE reaches the coverage area of a gNB, (e.g. gNB1 in Fig. 2), the UE begins to report geo-localised measurements to the network. 3GPP standards provide different ways to determine the geographic position and/or velocity of the UE based on measuring radio signals [1]. These positioning methods include network-assisted GNSS (Global Navigation Satellite System) methods, Time Difference of Arrival (TDoA) based on 4G Long Term Evolution (LTE) or 5G New Radio (NR) signals, Wireless Local Area Network (WLAN) positioning, Bluetooth positioning, etc. Location measurements collected from UEs in RRC_connected state can be transmitted periodically to the network as part of the radio measurement reporting processes [28]. In turn, UEs in RRC_idle or RRC_inactive mode can log measurements and transmit them later on when the UE enters in RRC_connected state, e.g. using the MDT (Minimization of Drive Tests) feature [29]. The collected UEs location measurements associated to each gNB are sent to the OAM system. With this data, the OAM is able to build a historical list of UE trajectories inside the coverage region of each gNB. Then, the *Model training* process represented in Fig. 2 runs the clustering of these historical trajectories in order to learn and model the most representative UE trajectories (i.e. prototype trajectories) in each gNB and store them in a database in the OAM. With a specific periodicity, new collected UE trajectories may be included in the trajectory modelling process in order to periodically update the database of prototype trajectories.

Once the training is finalised, the obtained prototype trajectories, which constitute the learnt model, are transferred to each gNB in the AI/ML model deployment step. Then, the inference stage carries out the online mobility prediction of the UEs served by the gNB. During the inference, a UE reports location measurements to its serving gNB. Then, by comparing these location reports with the prototype trajectories, the *Model inference* determines the prototype trajectory that the UE will likely follow in the near future. This will allow predicting the future positions of this UE, the next gNB that the UE will visit and the estimated time for handover to this new gNB. Finally, the obtained predictions are used, in this example, for the optimization of the handover procedure. The proposed methodology can be easily adapted to the rest of solutions and use cases defined in [17]. It is worth mentioning that, although the proposed approach has been presented in the context of the 3GPP framework, it can be also aligned with other frameworks, such as Open-RAN (O-RAN) specifications [30]. In this case, the *Model training* can be run in the non-RT RIC (non-Real Time RAN Intelligent Controller) and the *Model inference* can be run in the near-RT RIC (near Real Time RAN Intelligent Controller) to assist the operation of the gNB-CU.

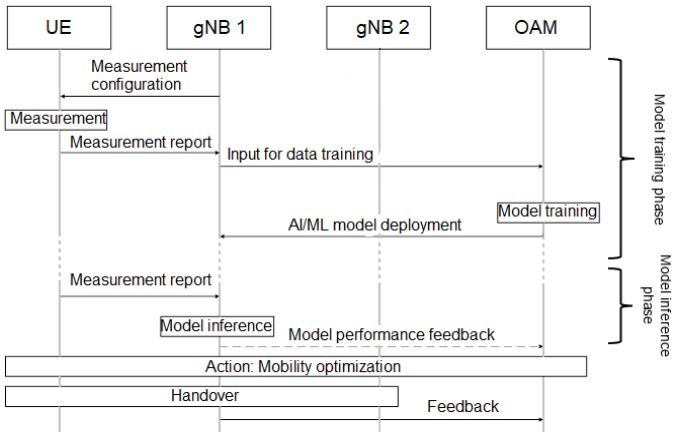


Fig. 2. Solution for AI/ML enabled mobility optimisation.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

IV. PROPOSED CLUSTERING AND PREDICTION METHODOLOGY

A. General overview

The proposed methodology consists of two main steps:

First, in the *Model training* process, the historical locations reported by the UEs and available in the OAM are pre-processed in order to build a set of historical UE trajectories for each cell. Then, in a cell-by-cell basis, a trajectory clustering process is done for all the UE trajectories of the same cell. After the clustering process, a list of prototype trajectories is obtained for each cell. This list of prototype trajectories provides a characterization of the most frequent and relevant trajectories followed by the UEs along each cell. Further details of the *Model training* based on a trajectory clustering are provided in section IV.B.

The *Model inference* process aims to provide an on-line prediction of the future locations of currently connected UEs. When a specific UE enters in the coverage region of a specific gNB, the UE initiates the process of reporting its current geographical location as it moves inside the cell. Each time a new UE location is reported, the current trajectory followed by the UE is compared with the different prototype trajectories of this cell in order to determine, with certain likelihood, which is the most likely trajectory that the UE will follow inside the cell. The details of the *Model inference* carried out for UE mobility prediction are provided in section IV.C.

B. Model training (UE trajectory clustering)

The proposed methodology for learning the prototype trajectories is illustrated in Fig. 3. It assumes a cellular system with B Base Stations (BS) or gNBs, each one handling a cell that provides coverage on a certain geographical area. First, the collected UE location measurements are processed in order to build a set of H_b trajectories followed by the UEs in the coverage region of each b-th BS with $b=1,\ldots,B$ during a time period of D days. Each UE trajectory is represented as a sequence of consecutive geographical locations reported by each UE every T seconds. Then, for each b-th BS, the $\boldsymbol{r_{h,b}}$ trajectory ($h=1,\ldots,H_b$) is defined as the concatenation of the $N_{h,b}$ coordinates of consecutive time instants denoted as $\boldsymbol{r_{h,b}}=[r_{h,b}(1),r_{h,b}(2),\ldots,r_{h,b}(N_{h,b})]$, where $r_{h,b}(i)$ corresponds to the two-dimensional (x,y) location i.e. $(r_{x,h,b}(i), r_{y,h,b}(i))$ of the UE at the i-th sample of the corresponding trajectory. In this process, for each of the H_b trajectories that belong to the b-th BS, the previous and the next serving BS identifier visited by the UE are also determined. In order to simplify the trajectory modelling process, all the H_b trajectories of the same b-th BS are pre-processed in order to have the same number of samples equal to $N_b=max(N_{h,b})$ with $h=1,\ldots,H_b$. This is done by interpolating samples in each of the H_b trajectories. In order to do this, first, the length of each $\boldsymbol{r_{h,b}}$ trajectory $d_{tot,h,b}$ is calculated as the summation of all the Euclidean distances between two consecutive samples for all the trajectory:

$$d_{tot,h,b} = \sum_{i=1}^{N_{h,b}-1} \sqrt{\left[r_{x,h,b}(i) - r_{x,h,b}(i+1)\right]^2 + \left[r_{y,h,b}(i) - r_{y,h,b}(i+1)\right]^2} \quad (1)$$

The location samples of the interpolated trajectory $\boldsymbol{r'_{h,b}}=[r'_{h,b}(1), r'_{h,b}(2), \ldots, r'_{h,b}(N_b)]$ are calculated in a way that the distance between two consecutive samples (i.e. $dist\left[r'_{h,b}(i), r'_{h,b}(i+1)\right]$ for all $i\epsilon[1,N_b\text{-}1]$) is constant and equal to $\Delta_{h,b}=d_{tot,h,b}/N_b$. Then, the trajectory clustering process is done for all the interpolated trajectories $\boldsymbol{r'_{h,b}}$ in a *BS-by-BS* basis. The proposed methodology supports different kind of clustering algorithms such as K-means [31], Self Organising Maps (SOM) [32], DBSCAN [33], etc.

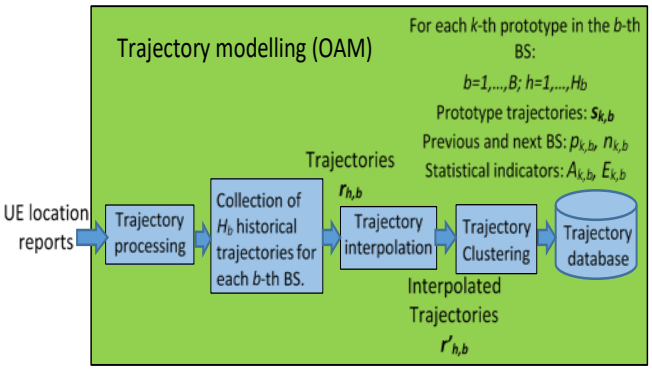


Fig. 3 Off-line procedure for learning mobility patterns.

In this paper, the clustering process is done according to the K-means methodology (see Algorithm 1) because of its simplicity, popularity, robustness and capability to manage large datasets with a relatively low computational cost. According to this, the H_b interpolated trajectories $\boldsymbol{r'_{h,b}}$ of a given b-th BS are grouped in K_b clusters in a way that trajectories of the same cluster are similar among them and different from the trajectories belonging to the rest of the clusters. In order to determine the most adequate value of K_b, the K-means process is run for different values of the number of clusters between K_{min} and K_{max}. Then, the selected value of K_b is the one that provides a clustering with the minimum Davies-Bouldin index [34]. This index takes into account how similar are all the trajectories that belong to the same cluster and how different are the obtained prototype trajectories of the different clusters. Low values of the Davies-Bouldin index reflect a better quality of the clustering process.

The result of the trajectory clustering for each b-th BS consists of a set of K_b prototype trajectories. The k-th prototype trajectory is represented as $\boldsymbol{s_{k,b}}=[s_{k,b}(1),s_{k,b}(2),\ldots,s_{k,b}(N_b)]$ where $s_{k,b}(i)$ is the i-th geographical location (x,y). $\boldsymbol{s_{k,b}}$ is determined as the centroid of all the trajectories that belong to the same cluster (i.e. each $s_{k,b}(i)$ is calculated as the average location of each of the positions $r'_{h,b}(i)$ for all the trajectories of the same cluster). In addition, several statistical indicators are also stored for each cluster:

- *Percentage of hits ($A_{k,b}$):* It is the percentage of trajectories associated to the b-th BS that belong to the k-th cluster (with $k=1,\ldots,K_b$). The prototype trajectories of clusters with a high percentage of hits will be more frequent and representative of the trajectories of each BS.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

• *Average Euclidean distance of the trajectories in a cluster ($E_{k,b}$):* It is a metric that captures the degree of similarity between trajectories of the same cluster with respect to the prototype trajectory $\boldsymbol{s_{k,b}}$ of the cluster. This is calculated as the average Euclidean distance between each trajectory of a cluster and the corresponding prototype trajectory of this cluster. A high value of $E_{k,b}$ reflects a higher dispersion in the cluster, meaning that the prototype trajectory is less representative of the clustered trajectories.

• *Previous and next serving BS identifier:* For each k-th prototype trajectory in each b-th BS, the previous $p_{k,b}$ and next $n_{k,b}$ serving BS identifier are calculated as the mode (i.e. most frequent value) of previous and next BS for all the trajectories that belong to the same cluster.

Algorithm 1. Trajectory clustering based on K-means.

1 **For** $K_b=K_{min}$ until $K_b=K_{max}$
2 #Run the clustering to obtain K_b clusters.
3 - Select randomly K_b out of the H_b trajectories. Each of the K_b trajectories represents an initial cluster with a centroid equal to the corresponding trajectory.
4 - Each remaining H_b-K_b trajectory is assigned to the closest cluster according to the Euclidean distance between the trajectory and the centroid.
5 **Do**
6 - Compute the new centroid of each of the K_b clusters.
7 - Reassign each of the H_b trajectories to the closest cluster according to the Euclidean distance between the trajectory and the centroid of each cluster.
8 **While** no changes observed in the clustering in two consecutive iterations.
9 **end**
10 Compute Davies-Bouldin index for the obtained clustering.
11 **end**
12 Select the value of K_b that provides the minimum Davies-Bouldin index.

C. Model inference (UE mobility prediction)

The proposed mobility prediction approach is run in a *UE-by-UE* basis and considers each new UE that enters in the coverage region of a certain b*-th BS. The trajectory prediction process is illustrated in Fig. 4. It starts with the collection of the geolocation coordinates reported by the UE to the b*-th BS in order to build the trajectory $\boldsymbol{u}$ that is currently being followed by this UE. It is assumed that whenever this UE enters in a region covered by the b*-th BS, the $\boldsymbol{u}$ vector is initialised as $\boldsymbol{u}=[u(1)]$ with the current UE location coordinates (x,y). After M UE location measurements the trajectory $\boldsymbol{u}=[u(1),u(2),\ldots,u(M)]$ is a vector composed by the concatenation of these M pairs of coordinates followed by the UE at consecutive time instants. The mobility prediction is run in the *b*-th* BS each time a new UE location measurement is collected and determines the likelihood that the $\boldsymbol{u}$ trajectory followed by the UE is one of the learnt prototype trajectories $\boldsymbol{s_{k,b*}}$ associated to the b*-th BS. The trajectory $\boldsymbol{u}$ and the prototype trajectories $\boldsymbol{s_{k,b*}}$ are interpolated in a way that the distances between two consecutive samples of each of these trajectories is equal to Δ, i.e. $dist[u(i),u(i+1)]=\Delta$ for all $i\epsilon[1,M\text{-}1]$ of the $\boldsymbol{u}$ trajectory and $dist[s_{k,b^*}(i),s_{k,b^*}(i+1)]=\Delta$ for all $i\epsilon[1,N_{b^*}\text{-}1]$ in all the $\boldsymbol{s_{k,b*}}$ trajectories. The result of this interpolation is the interpolated trajectory $\boldsymbol{u'}$ consisting of M' samples $\boldsymbol{u'}=[u'(1),u'(2),\ldots,u'(M')]$ and the interpolated prototype trajectories $\boldsymbol{s'_{k,b*}}=[s'_{k,b*}(1),\ s'_{k,b*}(2),\ldots,s'_{k,b*}(N'_{b*})]$ each one consisting of N'_{b*} samples. After this interpolation, the trajectory prediction process aims to identify the $\boldsymbol{s'_{k,b*}}$ prototype trajectory with highest similarity to the $\boldsymbol{u'}$ trajectory followed by the UE according to Algorithm 2.

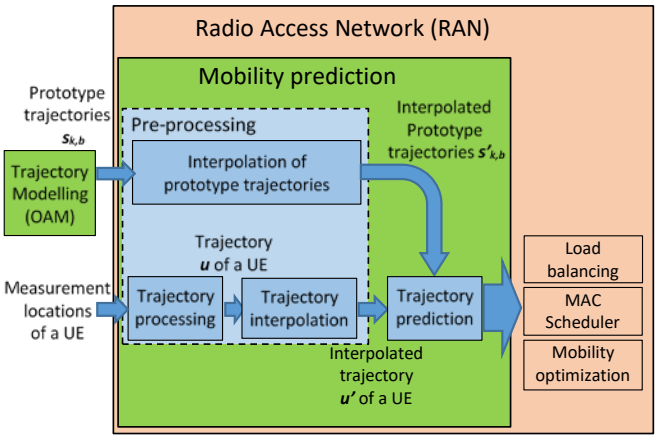


Fig. 4. On-line trajectory prediction

Without loss of generality, let us consider that the dimension of $\boldsymbol{u'}$ is lower than the number of elements of the prototype trajectories $\boldsymbol{s'_{k,b*}}$ (i.e. $M'\leq N'_{b*}$). This reflects that, in case that the UE was following a specific prototype trajectory, the actual location of the UE is somewhere within this prototype trajectory. In order to assess the similarity between $\boldsymbol{u'}$ and $\boldsymbol{s'_{k,b*}}$, the methodology considers all the α-*th* possible portions of M' consecutive elements of the vectors $\boldsymbol{s'_{k,b*}}$ ($k=1,\ldots,K_b$) as defined in line 3 in Algorithm 2. Then, the squared Euclidean distance $d_{u',s'_{k,b^*}}(\alpha)$ between the α-th portion of $\boldsymbol{s'_{k,b*}}$ and the UE trajectory $\boldsymbol{u'}$ is calculated (see lines 4-7 in Algorithm 2). The similarity between $\boldsymbol{u'}$ and $\boldsymbol{s'_{k,b*}}$ is denoted as m_{k,b^*} and it is computed as the minimum value of the square Euclidean distance between $\boldsymbol{u'}$ and the all the possible α-th portions of the prototype trajectory $\boldsymbol{s'_{k,b*}}$. A low value of m_{k,b^*} indicates that the trajectory $\boldsymbol{u}$' is very similar to some portion of vector $\boldsymbol{s'_{k,b*}}$. If the lowest value of m_{k,b^*} for all the K_{b^*} prototype trajectories is higher than a similarity threshold Th_S (see line 11 in Algorithm 2), then, the methodology identifies that the current trajectory is not enough similar to any prototype trajectory and, as a consequence, it does not correspond to any of them. Otherwise, the likelihood L_{k,b^*} that the UE is following the prototype trajectory $\boldsymbol{s'_{k,b*}}$ is determined according to the m_{k,b^*} values for all the K_{b^*} prototype trajectories. The percentage of hits A_{k,b^*} of each prototype trajectory is also considered in the likelihood calculation (see line 15 in Algorithm 2) in order to provide higher likelihood to prototype trajectories with higher representativeness. A high value of L_{k,b^*} reflects that the UE is following a trajectory very similar to a portion of $\boldsymbol{s_{k,b*}}$. Therefore, $\boldsymbol{s_{k,b*}}$ provides information about the positions that the UE may likely follow in the future. In case that the highest likelihood is obtained for the prototype trajectory k*, and this value is higher than a specific threshold (i.e. $L_{k^*,b^*}>Th_L$), then,

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

the methodology assumes that the UE is following this prototype trajectory $\boldsymbol{s_{k^*,b^*}}$ and the next BS identifier is n_{k^*,b^*}. In this case, according to the current UE location, it is possible to determine the remaining distance for the UE to reach the cell edge following the identified $\boldsymbol{s_{k^*,b^*}}$ prototype trajectory. The remaining time for handover t_{HO} can be estimated according to the remaining distance for the UE to reach the cell edge and an estimation of the average UE speed following the $\boldsymbol{u}$ trajectory. The obtained result can be used for e.g. optimizing the handover process. For example, in case that the estimated time for handover is lower than a specific threshold Th_{HO} (i.e. $t_{HO}<Th_{HO}$) a resource reservation may be done at the next n_{k^*,b^*} BS to guarantee service continuity to the UE. In case that the highest likelihood is obtained in multiple prototype trajectories, the algorithm selects these trajectories and groups those that follow the same path from the current UE location until the end of the trajectory. Then, for each g-th group of prototype trajectories, a likelihood L_{g,b^*} is calculated as the summation of the likelihood for all the prototype trajectories that belong to the same group. In case that the group g with the highest likelihood fulfils $L_{g,b^*}>Th_L$, the methodology predicts the UE future trajectory, the next BS and provides an estimation of the time to reach the cell edge in a similar way as before (see lines 31-34 in Algorithm 2).

Algorithm 2. UE mobility prediction algorithm

1 **For** $k=1$ until $k=K_{b^*}$
2 ***For*** $\alpha=0$ *until* $\alpha=(N'_{b^*})-M'$
3 Define the α-th portion of $\boldsymbol{s'_{k,b^*}}$ as vector $[s'_{k,b^*}(1+\alpha),\ldots, s'_{k,b^*}(M'+\alpha)]$
4 Determine:
5 $d_{u',s'_{k,b^*}}(\alpha)=\sum_{m=1}^{M'}\left[u'_x(m)-s_x{}'_{k,b^*}(m+\alpha)\right]^2+\left[u'_y(m)-s_y{}'_{k,b^*}(m+\alpha)\right]^2$
6 where (u'_x,u'_y) and $(s_x{}'_{k,b^*}, s_y{}'_{k,b^*})$ represent the (x,y) coordinates
7 of each $\boldsymbol{u'}$ and $\boldsymbol{s'_{k,b^*}}$ element.
8 Determine
$$m_{k,b^*}=\min_{\alpha}\left[d_{u',s'_{k,b^*}}(\alpha)\right]$$
9 **end**
10 **end**
11 **if** $\min[m_{k,b^*}]>Th_S$
12 The UE trajectory does not match with any prototype
13 **else**
14 **For** $k=1$ until $k=K_{b^*}$
15 Determine $L_{k,b^*}=\frac{A_{k,b^*}/m_{k,b^*}}{\sum_{k=1}^{K_{b^*}}(A_{k,b^*}/m_{k,b^*})}$
16 **end**
17 **If** there is only one prototype trajectory k^* with the highest likelihood
18 **If** $L_{k^*,b^*}>Th_L$:
19 The methodology assumes that the UE trajectory corresponds to
20 the k^*-th prototype trajectory.
21 The methodology assumes that the next BS is n_{k^*,b^*}
22 Estimate the time for handover t_{HO} (i.e. time to reach the cell
23 edge).
24 **If** $t_{HO}<Th_{HO}$ reserve resources in the n_{k^*,b^*} BS.
25 **Else**: The UE trajectory cannot be predicted with enough likelihood
26 **end**
27 **Else**: #The highest likelihood is obtained for several prototypes
28 Group the prototypes that follow the same path from the current UE
29 location until the end of the prototype. Determine the aggregated
30 likelihood L_{g,b^*} for each g-th group of prototypes.
31 **If** $L_{g,b^*}>Th_L$ for a group of prototypes:
32 Determine the next BS and estimate t_{HO} (i.e. time to reach the
33 cell edge).
34 **If** $t_{HO}<Th_{HO}$ reserve resources in the identified next BS.
35 **Else**: The UE trajectory cannot be predicted with enough likelihood
36 **end**
37 **end**
38 **If** the UE trajectory cannot be predicted with enough likelihood.
39 #Check if, at least, it is possible to determine the next BS.
40 **For** each neighbour BS β in the list of neighbours of b^*
41 Determine
$$L_{\beta}=\sum_{k\ with\ n_{k,b^*}=\beta}L_{k,b^*}$$
42 Determine the β^*-th neighbour with the highest likelihood
43 **If** $L_{\beta^*}>Th_L$:
44 The methodology assumes that next BS for this UE is the
45 β^*-th neighbour BS
46 Estimation of minimum t_{HO_min} time for handover for all the
47 prototype trajectories with $n_{k,b^*}=\beta^*$
48 **If** $t_{HO\_min}<Th_{HO}$ reserve resources in the β^* BS.
49 **Else**: the next BS cannot be predicted with enough likelihood.
50 **end**
51 **end**
52 **end**
53 **end**

Finally, in case that no prototype trajectory can be predicted with enough likelihood, the methodology aims to, at least, predict the next BS. This is done by calculating the likelihood L_β that the next BS is the β-th BS, being β the identifier of one of the neighbours of the b^*-th BS. This is done by adding the likelihood L_{k,b^*} of all the prototype trajectories of the b^*-th BS that have the β-th BS as the next serving BS identifier (i.e. $n_{k,b^*}=\beta$) as shown in line 41 in Algorithm 2. Then, the β^*-th BS with the highest likelihood is determined. In case that the likelihood L_{β^*} is higher than threshold Th_L, then, the methodology assumes that the next BS for this UE is the β^*-th BS. In a similar way as before, in order to estimate the time for handover, the methodology determines the time to reach the cell edge following all the possible prototype trajectories of the b^*-th BS that have the β^*-th BS as the next serving BS identifier (i.e. $n_{k,b^*}=\beta^*$) and takes the minimum value t_{HO_min}. This term represents an estimation of the time to handover to the next serving BS identifier β^* following the prototype trajectory that needs the lowest time to reach this β^* neighbour BS. Although the prototype trajectory is not identified, the prediction of the next serving BS can be used for e.g. optimising the handover process. Finally, in case that $L_{\beta^*}<Th_L$, the next BS cannot be predicted with enough likelihood for this UE.

V. Scenario Description

The proposed trajectory clustering and prediction methodology has been tested in a real-life mobility scenario using the dataset [18] that includes mobility traces for more than 50000 vehicles moving around inside the city of Cologne during 24 hours. The considered parameters of the algorithm are summarised in Table 1. The region under study includes a total of $B=245$ BSs as shown in Fig. 5. The Voronoi tessellation [35], represented with orange lines in Fig. 5, has been used to determine the coverage area of a BS, where the Voronoi region of a BS is composed by the geographical locations that have this BS as the closest one. After processing the trajectories inside each BS for the available dataset, approximately 5.46 million

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

trajectories have been obtained (5.02 million have been used in the clustering process and the remaining 442295 have been used for testing the prediction methodology). The *K*-means trajectory clustering process is repeated with different values of the number of clusters from $K_{min}=2$ to $K_{max}=120$. As a result of the clustering process for each BS in the whole scenario, a total number of 18506 prototype trajectories have been obtained.

TABLE 1. CONSIDERED PARAMETERS

Time period of measurements	*D=1* day (i.e. 24hours)
Time between geographical samples	*T=1* second
Number of BSs	*B=245* BS
Number of trajectories used for training	*5.02* million trajectories
Number of trajectories used for prediction	*442295* trajectories.
Minimum number of clusters	$K_{min}=2$
Maximum number of clusters	$K_{max}=120$
Distance between two consecutive samples of the interpolated trajectory	*Δ=5* meters
Similarity threshold	$Th_S=20$ meters
Likelihood threshold (considered values)	$Th_L=0.5$, *0.7* and *0.9*

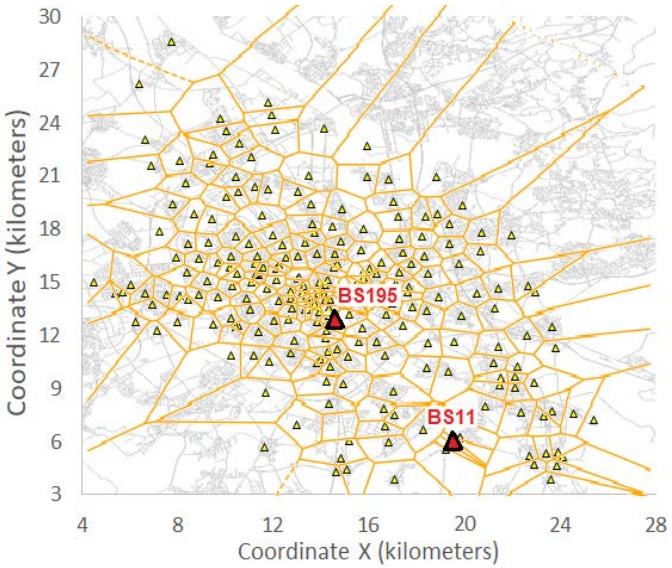


Fig. 5. Considered region in the city of Cologne (Germany)

The performance of the proposed methodology has been evaluated in terms of its capability to predict the future UE trajectory and the next BS. The following Key Performance Indicators (KPI) have been evaluated for assessing the UE trajectory prediction:

- *Prototype prediction accuracy*: This is the percentage of UE trajectories correctly predicted with a likelihood higher than Th_L before the UE trajectory inside the current cell has been completed.
- *Elapsed time before prediction*: This is the elapsed time between the instant of time when the UE begins its trajectory in the current cell and the time when the trajectory is correctly predicted with a likelihood that remains higher than Th_L during the rest of UE the trajectory.
- *Travelled distance before prediction*: This is the distance travelled by the UE during the elapsed time before prediction.

In turn, concerning the evaluation of the performance of the next BS prediction, the following KPIs have been considered:

- *Next BS prediction accuracy*: This is the percentage of UE trajectories in which the next BS visited by the UE is correctly predicted with a likelihood higher than Th_L before the UE trajectory inside the current cell has been completed.
- *Prediction anticipation time*: This is the difference between the instant of time when the next BS is correctly predicted with a likelihood that remains higher than Th_L and the time when the UE finishes its trajectory inside the current cell.
- *Prediction anticipation distance*: This is the distance travelled by the UE during the prediction anticipation time.

VI. RESULTS

As shown in Fig. 5, the considered scenario contains multiple types of cells. Some of them are located in suburban areas and provide coverage to regions in which UEs follow very simple trajectories along a main road and may change their direction in a reduced number of crossroads (e.g. BS11). Moreover, this kind of BSs usually have a small number of neighbor BSs. For these reasons, a high predictability is expected in this kind of cells. Other BSs are located in urban regions and provide coverage in areas with a higher number of streets and crossroads (e.g. BS195 located in the city center as shown in Fig. 5). As a consequence, higher number of possible UE trajectories inside the cell are observed. In addition, these urban regions usually have a higher density of BSs leading to a higher number of neighbor BSs. For these reasons, a lower predictability is expected in this kind of cells. In order to illustrate the performance of the proposed methodology in such different scenarios, section A shows the overall performance obtained in a suburban BS (i.e. BS11) and the analysis of some specific trajectories inside this cell. A similar study is presented in Section B for BS195 that represents a cell located in an urban area. Finally, Section C provides an overall evaluation for all the *B=245* BSs.

A. Analysis of a BS in an interurban area

A.1. Trajectory clustering.

Fig. 6 shows the collected UE locations in the Voronoi region of BS11. The borders between different Voronoi regions are represented in orange. A total of H_{11}=4487 trajectories have been used in the clustering process, while 460 trajectories have been used for the evaluation of the prediction methodology. After running the clustering process in BS11, the number of clusters that minimises the Davies-Bouldin index is $K_{11}=10$ clusters. Fig. 7 presents the obtained prototype trajectories, while Table 2 shows the number of hits $A_{k,11}$, the Euclidean distance $E_{k,11}$ and the previous $p_{k,11}$ and next $n_{k,11}$ serving BS identifier for each of the obtained clusters. As shown in Fig. 7, the prototype trajectories represent different UE movements along the main streets inside BS11. The most frequent observed trajectories correspond to prototype trajectories 0 and 1 ($A_{0,11}=49.36\%$ and $A_{1,11}=43.86\%$). Note also that, as shown in Table 2, some prototype trajectories represent UEs that come from a neighbour cell and, after going through BS11, they move

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

to a new neighbour cell. However, in other identified clusters, the prototype trajectory represents UEs that begin or end its movement at this BS11. This is represented as "--" in Table 2. In particular, prototype trajectory 6 represents UEs that begin their movement in BS11 and then move to BS206, while, prototype trajectories 4, 7 and 9 represent UEs that come from BS65 or BS206 and finish their movement at BS11 (see Fig. 7).

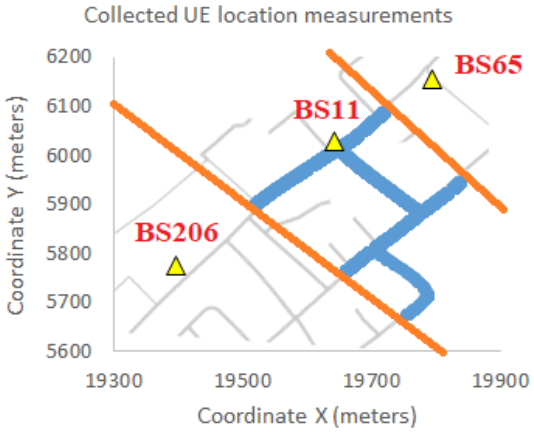


Fig. 6. Collected UE locations in the Voronoi region of BS11

A.2. Prediction of different UE trajectories in BS11.

In order to illustrate the performance of the prediction methodology, the result of the prediction of different UE trajectories is presented in this section. In addition, the benefit of including the percentage of hits of each prototype trajectory in the calculation of the likelihood is compared to the case in which the percentage of hits is not considered as it was done in [36]. Fig. 8a shows a map with the trajectory of a UE that comes from BS65, moves in the region of BS11 in the south-west direction and finally reaches the region of BS206, while Fig. 8b and 8c present the evolution of the likelihood of the different prototype trajectories as the UE is moving. In this example, the likelihood threshold is set to $Th_L=0.7$. As shown in Fig. 8b, when the percentage of hits is not considered in the likelihood calculation, every time a new UE trajectory sample is collected, the methodology determines a set of candidate prototype trajectories that contain the current UE trajectory and provides the same likelihood for all these candidate prototypes. As a consequence, at the beginning, the methodology is not able to provide a high likelihood $L_{k,11}$ for any of the prototype trajectories. However, as new location samples are obtained while the UE is moving in the south-west direction, some prototype trajectories are discarded (see Fig. 8b). After six UE trajectory location samples (i.e. six seconds of trajectory) the methodology is able to determine that the UE trajectory corresponds to prototype trajectory 1 (i.e. cluster 1). Then, the methodology identifies that the next BS will be BS206 with an anticipation of 3 seconds. It is worth noting that, in this example, changing the likelihood threshold to e.g. $Th_L=0.5$ or $Th_L=0.9$ has a marginal impact in the anticipation time (see Fig 8.b). In case that the percentage of hits is considered in the calculation of the likelihood, the methodology identifies that the

TABLE 2. STATISTICS FOR THE OBTAINED CLUSTERS IN BS11

Cluster	Number (and percentage) of hits ($A_{k,11}$)	Average distance to the centroid ($E_{k,11}$)	Previous BS ($p_{k,11}$)	Next BS ($n_{k,11}$)
Cluster 0	2215 (49.36%)	6.33	206	65
Cluster 1	1968 (43.86%)	6.76	65	206
Cluster 2	146 (3.25%)	7.71	65	65
Cluster 3	44 (0.98%)	5.14	206	206
Cluster 4	2 (0.04%)	10.94	65	--
Cluster 5	1 (0.02%)	0	206	206
Cluster 6	51 (1.13%)	2.46	--	206
Cluster 7	10 (0.22%)	18.49	206	--
Cluster 8	29 (0.64%)	5.26	206	65
Cluster 9	21 (0.46%)	6.90	65	--

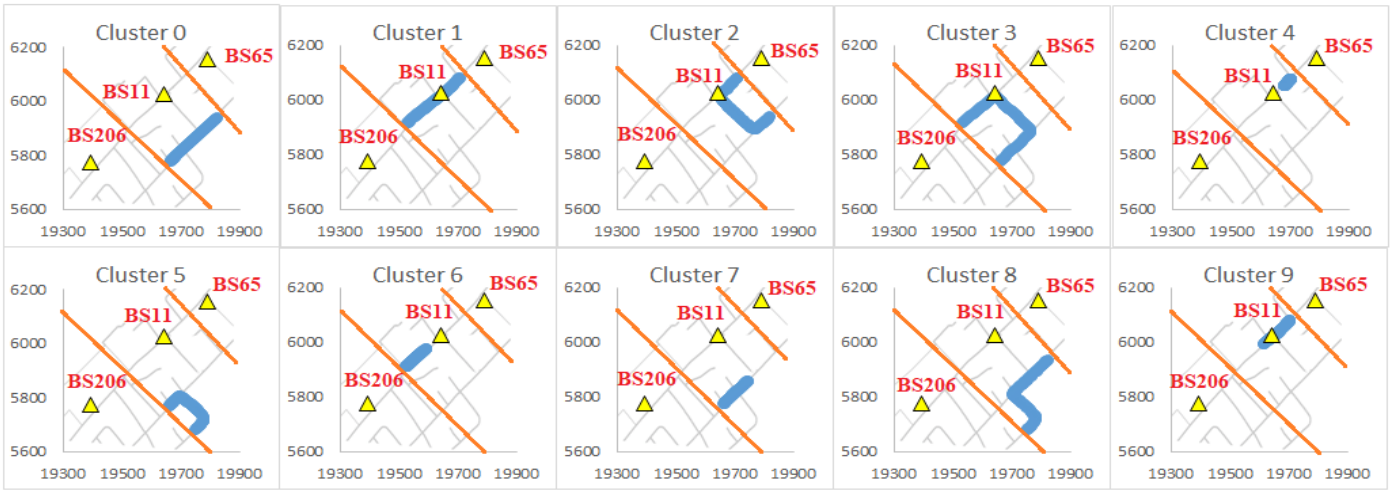
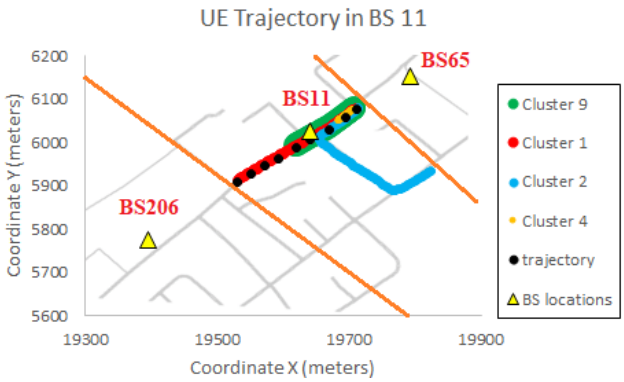


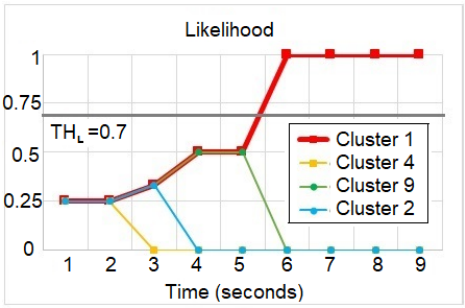
Fig. 7. Prototype trajectory for each cluster.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

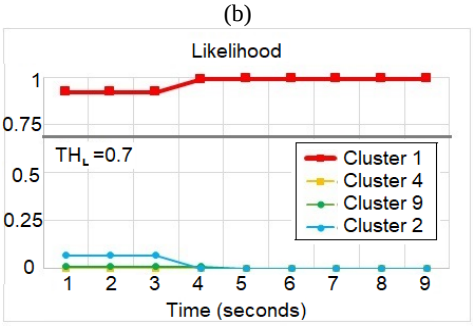
UE trajectory corresponds to prototype trajectory 1 from the very beginning of the UE trajectory (see Fig. 8c) because cluster 1 has a very high percentage of hits, as shown in Table 2.



(a)



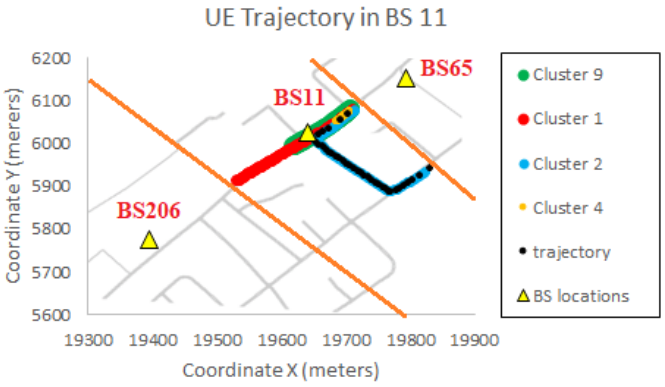
(b)



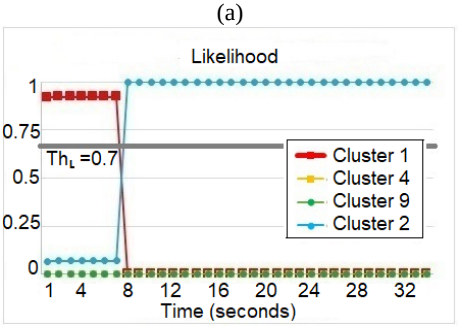
(c)

Fig. 8. a) Map of BS11, b) Likelihood according to [36], c) Likelihood with the proposed approach.

Fig. 9 presents another example for a different UE trajectory in BS11. In this case, the UE comes from BS65 and moves in the south-west direction, then, it turns left and goes in the south-east direction and finally it turns left again and goes in the north-east direction (see Fig. 9a). As shown in Fig. 9b, initially, the prediction methodology provides a high likelihood for prototype trajectory 1 since it has a very high percentage of hits (see Table 2). However, after 9 seconds (i.e. when the UE turns left for the first time), prototype trajectory 2 is predicted with a very high likelihood since the rest of trajectories are discarded. As shown, the prototype trajectory and the next BS can be correctly predicted after 9 seconds of UE movement (i.e. with an anticipation of 24 seconds). In this example, selecting a likelihood threshold $Th_L=0.5$ or $Th_L=0.9$ has no impact on the anticipation time.



(a)



(b)

Fig. 9. a) Map of BS11, b) Likelihood with the proposed approach.

A.3. Evaluation of global prediction statistics.

Table 3 shows the results of the UE trajectory and the next BS prediction obtained after evaluating the proposed methodology for the 460 trajectories available in BS11. The percentage of hits of each prototype is considered for the likelihood calculation. As shown in Table 3, both the prototype trajectory and the next BS visited by the UE are correctly predicted for almost all the UE trajectories. If the likelihood threshold is set to $Th_L=0.7$, the average trajectory time needed to correctly predict the prototype trajectory is around 2.31 seconds that corresponds to an average distance of 59.03 meters travelled ty the UEs (see Table 3). This time is relatively small when compared to the average dwell time of the users in BS11 that is around 10.10 seconds. Note also that the next BS prediction is done with a time anticipation of 7.79 seconds. This time is large enough to e.g. perform a reservation process in the identified neighbor BS to facilitate the handover process and guarantee service continuity.

TABLE 3. PROTOTYPE TRAJECTORY AND NEXT BS PREDICTION IN BS11

Statistics of prototype trajectory prediction	Th_L=0.5	Th_L=0.7	Th_L=0.9
Trajectory prediction accuracy (%)	97.82	97.82	97.82
Average elapsed time before prediction (seconds)	2.31	2.31	2.43
Average distance before prediction (meters)	59.03	59.03	60.70
Statistics of the next BS prediction	Th_L=0.5	Th_L=0.7	Th_L=0.9
Next BS prediction accuracy (%)	97.82	97.82	97.82
Average prediction anticipation time (seconds)	7.79	7.79	7.67
Average prediction anticipation distance (meters)	200.45	200.45	198.77

The impact of the likelihood threshold is also evaluated in Table 3. It is worth noting that, in this BS11, setting a higher

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

value of the likelihood threshold (e.g. Th_L=*0.9*) leads to a slightly higher average time needed to predict correctly the UE trajectory and a slightly lower anticipation time for the next BS prediction. The reason is that, when setting a higher value of the likelihood threshold, a higher likelihood is required to assume that the predictions done by the methodology are correct.

B. Analysis of a BS in an urban area

This section analyses the proposed methodology in BS195. As shown in Fig. 10, this BS covers a larger number of streets and crossroads than BS11.

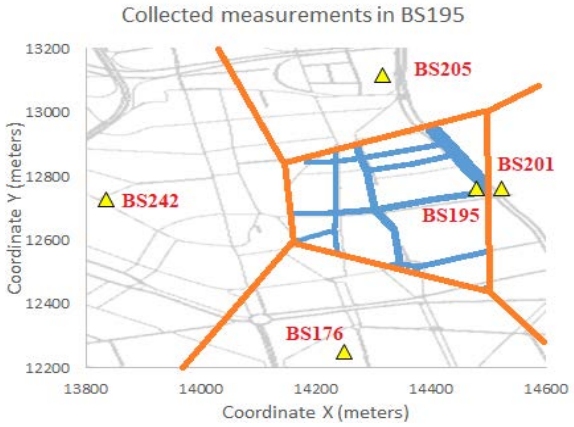


Fig. 10. Collected UE locations in BS195

A total of H_{195}=22954 trajectories have been used in the clustering process while the remaining 2320 trajectories have been used for the evaluation of the prediction methodology. After running the clustering process, the optimum number of clusters is K_{195}=*87* which provides a Davies-Bouldin index of 0.28. Due to the large amount of possible and diverse prototype trajectories identified in BS195 the trajectory and next BS predictions become more challenging. In order to illustrate this, Fig. 11 shows the result of the prediction of the trajectory represented in black colour for a UE that comes from BS176, moves inside the region of BS195 and finally moves to BS205. As shown in Fig. 11a, initially the methodology is able to identify a set of eight candidate prototype trajectories with different associated likelihoods (see Fig. 11b). As the UE is moving and new UE location samples are processed, some prototype trajectories are discarded. In particular, at time=12 seconds (i.e. when the UE moving to north turns right to the east) prototype trajectories 42 and 78 are discarded. At time=21 seconds (i.e. when the UE moving to the east direction turns left to the north) prototype trajectory 14 is also discarded. In any case, at time=21 seconds, no prototype trajectory has been identified yet with a likelihood higher than Th_L=*0.7*. However, since prototypes 51, 60 and 67 have BS205 as the next BS and the summation of these likelihoods is higher than Th_L=*0.7*, the methodology can predict BS205 as the next BS visited by the UE (see Fig. 11c).

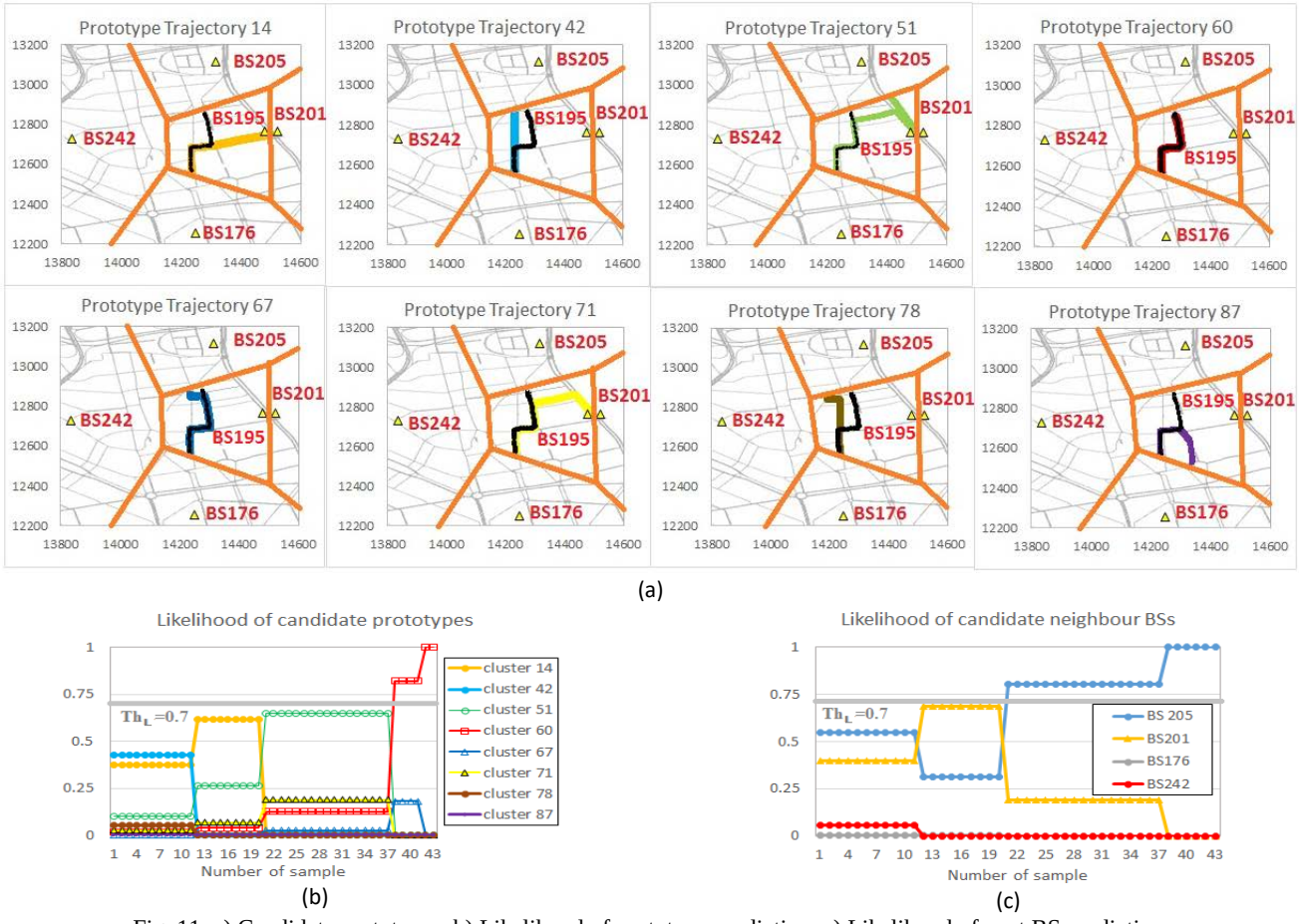


Fig. 11. a) Candidate prototypes, b) Likelihood of prototype prediction, c) Likelihood of next BS prediction.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

Concerning the trajectory prediction (see Figure 11b), at time=38 seconds, prototypes 51 and 71 are discarded and the methodology predicts prototype trajectory 60 with a likelihood higher than $Th_L=0.7$ with an anticipation of 5 seconds before the end of the UE trajectory. In case the likelihood threshold is set to $Th_L=0.9$, the methodology would not provide any prediction until time=38 seconds and time=41 seconds when BS205 and prototype trajectory 60 would be correctly predicted, respectively (see Figures 11b and 11c). In turn, in case that $Th_L=0.5$, the methodology would initially make several wrong predictions before making the correct one (i.e. at time=12seconds the methodology would wrongly predict prototype 14, at time=21seconds prototype 51 would be wrongly predicted and, finally, at time=38 seconds prototype 60 would be correctly predicted, see Figure 11b).

Table 4 shows the global prediction results obtained in BS195. As seen, the prediction accuracy for the next BS prediction is higher than the accuracy for the UE prototype trajectory prediction. The rationale of this is that, in some situations, the methodology is not able to determine the UE prototype trajectory among a set of possible candidate prototypes, but the next BS can be correctly predicted since most of these candidate trajectories have the same next BS. Note also that the prediction accuracies are lower in BS195 than in the results obtained in BS11 (see Tables 3 and 4). If the likelihood threshold is set to $Th_L=0.7$, the required time to predict the prototype trajectory in BS195 (i.e. 11.47 seconds) is rather high when compared to the average dwell time (i.e. 15.51seconds) which indicates that, on average, it is necessary to observe around 74% of a trajectory to do a reliable prototype trajectory prediction. The selected likelihood threshold has a clear impact in the obtained results. As shown in Table 4, a high likelihood threshold leads to a lower prediction accuracy, a higher required time for the UE trajectory prediction and a lower next BS anticipation time.

TABLE 4. PROTOTYPE TRAJECTORY AND NEXT BS PREDICTION IN BS195

Statistics of prototype trajectory prediction	Th_L=0.5	Th_L=0.7	Th_L=0.9
Trajectory prediction accuracy (%)	86.12	78.06	58.18
Average elapsed time before prediction (seconds)	4.60	11.47	12.89
Average distance before prediction (meters)	130.33	155.75	162.65
Statistics of the next BS prediction	Th_L=0.5	Th_L=0.7	Th_L=0.9
Next BS prediction accuracy	98.01	89.18	83.75
Average prediction anticipation time (seconds)	7.20	5.74	4.77
Average prediction anticipation distance (meters)	123.88	107.15	94.61

C. Global analysis of the prediction methodology

This section evaluates the global results when running the proposed methodology in the whole scenario consisting of $B=245$ BS. After testing the prediction methodology with a total number of 442295 UE trajectories, the percentage of correct UE trajectory predictions and next BS predictions have been determined for each BS. Fig. 12 and Fig. 13 present the Cumulative Distribution Function (CDF) of these two metrics for all the BSs in the scenario. The obtained prediction performance is highly dependent on the considered BS. As shown in Fig. 12, for the case of setting $Th_L=0.7$, the prototype trajectory is correctly predicted for all the UE trajectories in more than 25% of the BSs. However, there are some BSs in which the trajectory prediction accuracy is much lower (e.g. in 10% of the BSs, the prototype trajectory is correctly predicted for less than 62% of the trajectories). Note also in Fig. 13 that the next BS prediction accuracy is considerably higher than the UE trajectory prediction accuracy (e.g. in more than 55% of the BSs the next BS is correctly predicted for all the trajectories when considering $Th_L=0.7$).

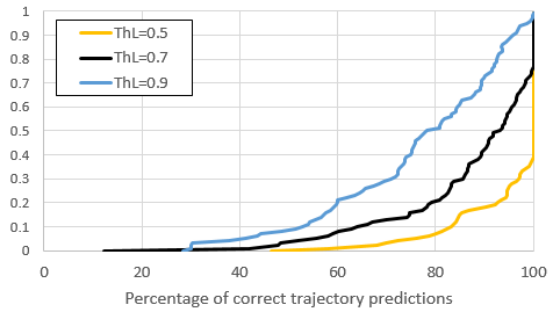


Fig. 12. CDF of the percentage of UE trajectory correct predictions

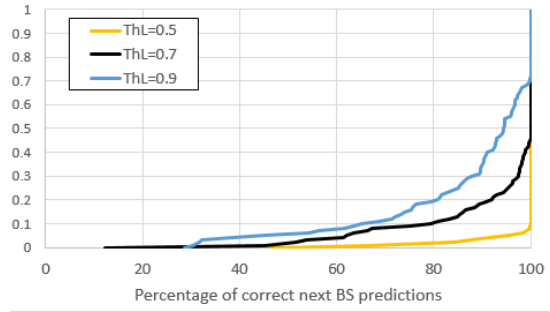


Fig. 13. CDF of the percentage of next BS correct predictions

Fig. 14 presents the CDF of the average anticipation time for next BS prediction. As shown, this metric is also highly dependent on the considered BS (e.g. in some BSs the average time anticipation is more than 50 seconds while there are a few BSs in which the average time anticipation is only 1 second). For the case of $Th_L=0.7$, the next BS can be predicted with an average time anticipation higher than 3 seconds for 99% of the BSs and higher than 10 seconds for 92% of the BSs (see Fig. 14). The impact of the likelihood threshold in the anticipation time is also evaluated in Fig. 14. For 99% of the BSs, an anticipation time higher than 4.45 seconds and 1.5 seconds is obtained for $Th_L=0.5$ and $Th_L=0.9$, respectively. Finally, the percentage of the UE trajectory required for a correct trajectory prediction is evaluated. This is calculated dividing the elapsed time before prediction by the average dwell time of each UE trajectory. This percentage is calculated and averaged for all the trajectories of a specific BS. Then, Fig. 15 presents the CDF of this percentage for all the BSs in the scenario. As shown in Fig. 15, when $Th_L=0.7$, in 10% of the BSs, the percentage of the UE trajectory required for a correct prediction is less than 27%. However, there are a 10% of BSs in which the required percentage of UE trajectory is higher than 82%.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3388554

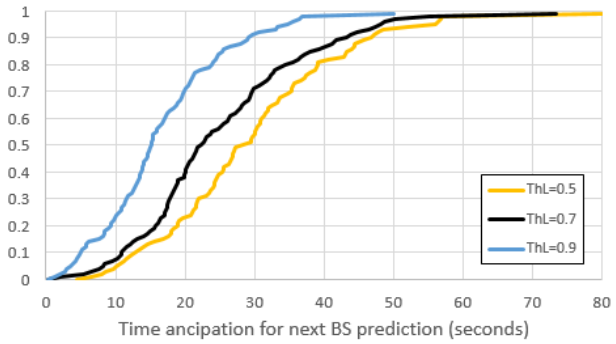


Fig. 14. CDF of the time anticipation of the next BS prediction

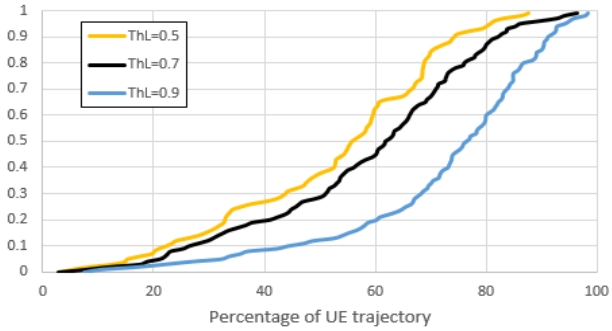


Fig. 15. CDF of the percentage of UE trajectory required for a correct trajectory prediction

D. Practical implementation considerations

This subsection provides a description of potential limitations of the proposed methodology and practical implementation considerations:

1. As described in section VI.C, the performance of the proposed prediction methodology is highly dependent on the considered BS. In some BSs that cover regions with a few streets/roads, the UE mobility may be quite predictable. However, in other BSs covering regions with a larger number of streets, the number of possible prototype trajectories may be rather large making the UE mobility prediction much more challenging. It would be recommended to disable the prediction methodology for these BSs with a relatively low prediction accuracy.

2. The proposed methodology requires a rich and a large database of UE trajectories in order to capture the relevant prototype trajectories inside each cell. In some specific cases, it may happen that a current UE trajectory inside a given cell does not match to any of the prototype trajectories currently stored for this cell, making it difficult the prediction for this specific trajectory. The proposed prediction methodology is capable of identifying such situations. Then, when this situation is detected, this new trajectory is included in the prototype trajectory database in the OAM as a new prototype trajectory to be considered for future UE mobility predictions in this cell.

3. UE mobility patterns are highly dependent on the time of the day, day of the week, etc. (e.g. some specific trajectories may happen more frequently at certain hours of the day or at the weekends). In order to provide a more accurate time/space UE mobility characterization and improve the prediction performance, the proposed methodology can be easily extended by storing a set of prototype trajectories at different time periods (different times of the day, different days of the week, etc.) for each cell.

4. The proposed methodology provides an estimation of the time when the UE will reach the next cell, based on the prediction of the prototype trajectory that the UE will follow inside the cell and its current average speed. A more precise estimation can be done if other context information is also included in the model (e.g. the traffic conditions in the area). As an example, information of the time spent by other UEs that have recently followed the same trajectory can be useful for a more accurate estimation of the time to reach the next cell.

VII. CONCLUSIONS

This paper has proposed a new methodology for UE mobility prediction based on the comparison of the recent UE locations in the current cell and a set of prototype trajectories previously obtained by means of a clustering process. The proposed methodology provides accurate predictions of the future trajectory followed by the UE in the current cell, the next cell visited by the UE and an estimation of the time when the UE will connect to this new cell. The proposed approach is aligned with the 3GPP functional framework for AI-enabled Radio Access Network (RAN) intelligence and the obtained predictions can be useful for improving several functionalities such as mobility optimization, load balancing and proactive resource allocation. The proposed methodology has been tested using a dataset that that includes mobility traces for more than 50000 vehicles inside the city of Cologne and considering a network deployment with hundreds of base stations. The obtained results are highly dependent on the selected likelihood threshold Th_L. A too high value of Th_L would lead to a too pessimistic evaluation of the prediction methodology since predictions are considered correct only when the likelihood is very high. In turn, setting a too low value of the likelihood threshold Th_L will lead to a too optimistic evaluation of the prediction methodology leading to higher percentage of correct predictions, a lower required trajectory time for prediction (or equivalently a higher anticipation time). If the likelihood threshold is set to $Th_L=0.7$, the prototype trajectory followed by a UE has been correctly predicted for 86.7% of UE trajectories while the prediction accuracy for the next BS visited by the UE is 94.4%. The obtained prediction results are also highly dependent on the considered BS. There are BSs that cover regions with a reduced number of streets and crossroads in which the UE trajectories can be highly predictable and the prediction accuracy can reach 100%. In other BSs, UEs may have multiple and diverse trajectories inside the BS resulting in lower prediction accuracy. Finally, the time required for UE trajectory prediction and the anticipation time of the next BS prediction has been evaluated. These metrics are also quite dependent on the considered BS. In general, if the likelihood threshold is set to $Th_L=0.7$, the next BS can be correctly predicted with an average time anticipation higher than 3

seconds for 99% of the BSs and higher than 10 seconds for 92% of the BSs.

REFERENCES

[1] 3GPP TS 38.305, "Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN", v16.7.0, Release 16, (2021-12).

[2] I. C-L, Y. Liu, S. Han, S. Wang, G. Liu, "On Big Data Analytics for Greener and Softer RAN", *IEEE Access*, August, 2015.

[3] R. Nisbet, G. Miner, K. Yale, *Handbook of Statistical Analysis and Data Mining Applications*, 2nd edition, Academic Press, November 2017.

[4] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," IEEE Signal Processing Magazine, vol. 31, no. 6, Nov. 2014.

[5] L. Huang, L. Lu and W. Hua, "A Survey on Next-Cell Prediction in Cellular Networks: Schemes and Applications," in *IEEE Access*, vol. 8, pp. 201468-201485, 2020, doi: 10.1109/ACCESS.2020.3036070.

[6] A. Hasbollah, S. H. S. Ariffin, and N. Fisal, ''Mobility prediction method for vehicular network using Markov chain,'' Jurnal Teknologi, vol. 78, nos. 6–2, Jun. 2016.

[7] H. Si, Y. Wang, J. Yuan, and X. Shan, ''Mobility prediction in cellular network using Hidden Markov Model,'' in Proc. 7th IEEE Consum. Commun. Netw. Conf. (CCNC), Jan. 2010, pp. 1–5.

[8] D. Katsaros, A. Nanopoulos, M. Karakaya, G. Yavas, O. Ulusoy, Y. Manolopoulos, "Clustering Mobile Trajectories for Resource Allocation in Mobile Environments", Advances in Intelligent Data Analysis. Lecture Notes in Computer Science, vol 2810. Springer, 2003.

[9] F. Meneghello, D. Cecchinato, M. Rossi, "Mobility Prediction via Sequential Learning for 5G Mobile Networks", 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2016.

[10] Y. Zhang, J. Hu, J. Dong, Y. Yuan, J. Zhou, J. Shi, "Location Prediction Model Based on Bayesian Network Theory", IEEE Global Communications Conference (GLOBECOM), Nov. 2009.

[11] W. Qi, Q. Song, S. Wang, Z. Liu, L. Guo, "Social Prediction-Based Handover in Collaborative Edge Computing Enabled Vehicular Networks", IEEE Transactions on Computational Social Systems, 2021.

[12] Z. Zhao, M. Karimzadeh, L. Pacheco, H. Santos, D. Rosario, T. Braun, E. Cerqueira, "Mobility Management With Transferable Reinforcement Learning Trajectory Prediction, IEEE Trans. On Network and Service Management, 2020.

[13] H. Farooq, A. Asghar, A. Imran, "Mobility Prediction Based on Proactive Dynamic Network Orchestration for Load Balancing with QoS Constraint (OPERA)", IEEE Transactions on Vehicular Technology, Vol. 69, No. 3, March 2020.

[14] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. Chan, and K. K. Leung, "Dynamic service migration in mobile edge computing based on Markov decision process" IEEE/ACM Transactions on Networking, vol. 27, no. 3, pp. 1272–1288, Jun. 2019.

[15] D. Ren, X. Gui, K. Zhang and J. Wu, "Mobility-Aware Traffic Offloading via Cooperative Coded Edge Caching," in *IEEE Access*, vol. 8, pp. 43427-43442, 2020, doi: 10.1109/ACCESS.2020.2977990.

[16] N. P. Kuruvatti,, S. B. Mallikarjun, S. C. Kusumapani, and H.D. Schotten. "Mobility Awareness in Cellular Networks to Support Service Continuity in Vehicular Users." In 2020 3rd International Conference on Information and Communications Technology (ICOIACT), pp. 431-435. IEEE, 2020.

[17] 3GPP TR 37.817 V17.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and NR. Study on enhancement for Data Collection for NR and EN-DC", Release 17, (2022-04).

[18] "Vehicular mobility trace of the city of Cologne (Germany)", available at http://kolntrace.project.citi-lab.fr/

[19] C. Yao, C. Yang, and Z. Xiong, "Energy-saving predictive resource planning and allocation" IEEE Trans. Commun., vol. 64, no. 12, pp. 5078–5095, Dec. 2016.

[20] J. Guo, C. Yang, and C.-L. I, "Exploiting future radio resources with end-to-end prediction by deep learning" IEEE Access, vol. 6, pp. 75729–75747, 2018

[21] H. Abou-zeid, H. S. Hassanein, R. Atawia, "Towards Mobility-Aware Predictive Radio Access: Modeling, Simulation and Evaluation in LTE Networks", International Conference on Modeling, analysis and simulation of wireless and mobile systems, MSWiM'14, September, 2014.

[22] M. Joud, M. García-Lozano, S. Ruiz, "User specific cell clustering to improve mobility robustness in 5G ultra-dense cellular networks", 14th Annual Conference on Wireless On-demand Network Systems and Services (WONS), 2018.

[23] J. Yang, C. Dai, and Z. Ding, ''A scheme of terminal mobility prediction of ultra dense network based on SVM,'' in Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA), Mar. 2017, pp. 837–842.

[24] Y. Gao, J. Chen, Z. Liu, L. Liu, N. Hu, "Deep Learning based Location Prediction with Multiple Features in Communication Network", IEEE Wireless Communications and Networking Conference (WCNC) 2021.

[25] Q. Liu, G. Chuai, J. Wang , J. Pan, "Proactive Mobility Management With Trajectory Prediction Based on Virtual Cells in Ultra-Dense Networks", IEEE Trans. on Vehicular Technology, Vol. 69, No. 8, August 2020.

[26] N. P. Kuruvatti, S.B. Mallikarjun, S. C. Kusumupani, H.D. Schotten, "Proactive Allocation of Radio Resources to Enhance Service Continuity in Cellular Networks", *IEEE International Symposium on Electrical and Electronics Engineering*, 2021.

[27] N. P. Kuruvatti, W. Zhou and H. D. Schotten, "Mobility Prediction of Diurnal Users for Enabling Context Aware Resource Allocation," *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1-5, doi: 10.1109/VTCSpring.2016.7504348.

[28] 3GPP TS 36331, "Radio Resource Control (RRC); Protocol Specification", v16.7, Release 16, (2021-12)

[29] 3GPP TS 37.320, "Radio measurement collection for Minimization of Drive Tests (MDT)", v16.7.0, Release 16, (2021-12).

[30] O-RAN: Towards an Open and Smart RAN, White Paper, October 2018.

[31] J. Han, M. Kamber, Data Mining Concepts and Techniques, 2nd edition, Elsevier, 2006.

[32] T. Kohonen, "Essentials of the Self-Organizing Map", Neural Networks, Elsevier, Vol. 37, 2013, pp. 52-65.

[33] E. Martin, H.P. Kriegel, J. Sander, X. Xu, S. Evangelos, H. Jiawei, U.M. Fayyad, "A density-based algorithm for discovering clusters in large spatial databases with noise". Second International Conference on Knowledge Discovery and Data Mining (KDD-96). August 1996.

[34] Davies, D. L.; Bouldin, D. W. "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol: PAMI-1, issue 2: pp: 224–227, April 1979. doi:10.1109/TPAMI.1979.4766909.

[35] Q.Du, V.Faber, M. Gunzburger, "Centroidal Voronoi Tesselations: Applications and Algorithms", SIAM Review 41 (4): 637–676, 1999.

[36] J. Sánchez-González, J. Pérez-Romero, R. Agustí, O. Sallent, "On Learning Mobility Patterns in Cellular Networks", 12th International Conference on Artificial Intelligence Applications and Innovations (AIAI 2016), 1st Workshop on 5G – Putting Intelligence to the Network Edge (5G-PINE), Thesalloniki, Greece, September 2016.

Juan Sánchez-González received the Telecommunication Engineer degree in May 2002 and the Doctor Engineer in Telecommunication degree in December 2007 from the Universitat Politècnica de Catalunya (UPC), Spain. He is currently associate professor at the UPC. He joined the Mobile Communications Research group of the Signal Theory and Communications department in 2002. His research interests are in the field of mobile and wireless communications systems, especially in radio resource and quality of service management and optimization of heterogeneous wireless networks. He has been involved in different national and european projects as well as in projects for private companies. He is currently working in the context

of the applicability of Artificial Intelligence/Machine Learning techniques for the optimization of heterogeneous wireless networks.

Oriol Sallent is a Professor at the Universitat Politècnica de Catalunya, Barcelona, Spain. He has participated in a wide range of European and national projects, with diverse responsibilities as a principal investigator, coordinator, and work package leader. He regularly serves as a consultant for a number of private companies. He has contributed to standardization bodies such as 3GPP, IEEE, and ETSI. He is the coauthor of 13 books and has authored or coauthored 300+ papers, mostly in high-impact IEEE journals and renowned international conferences. His research interests include 5G RAN (Radio Access Network) planning and management, artificial intelligence-based radio resource management, virtualisation of wireless networks, cognitive management cognitive radio networks and dynamic spectrum access and management, among others.

Jordi Pérez-Romero (Member, IEEE) received a degree in telecommunications engineering and a Ph.D. degree from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1997 and 2001, respectively. He is currently a Professor with the Department of Signal Theory and Communications of UPC. He is working in the field of wireless communication systems, with a particular focus on 5G and beyond cellular systems, including radio resource management and network optimization. He has been involved in different European projects with different responsibilities, such as researcher, work package leader, and project responsible, has participated in different projects for private companies and has contributed to the 3GPP and ETSI standardization bodies. He has coauthored more than 300 papers in international journals and conferences. He has also coauthored three books and has contributed to seven book chapters. He is an Associate Editor for the IEEE Vehicular Technology Magazine.