Localization of Mobile Devices in 5G Wireless Networks

Dandan Wang, Pantelis Monogioudis, Gurudutt Hosangadi, Anil Rao
Nokia
Murray Hill, NJ, USA
{*dandan.wang, pantelis.monogioudis, gurudutt.hosangadi, anil.rao*}*@nokia.com*

***Abstract*—As wireless network is evolving into 5G, tremendous data will be shared on the newly developed open source platforms. These data can be used in developing new service. Among which, Location information of mobile devices are exetremly useful. For example, the location information can be used to assist wireless operators to trouble shoot the network performance. It can also be used to provide some location assisted service. However, some of these devices may be designed for limited budget that do not have the capability of GPS. Furthermore, operators may not have access to the GPS information on the mobile devices. In this paper, we propose a novel machine learning based approach to estimate the location of the mobile devices based on the measurement data that mobiles reported during every call and session. Our proposed algorithm utilizes the advanced features of 5G wireless network, such as the beam information. Simulaiton shows that the proposed solution can achieve 10 m accuary for outdoor mobiles. And the proposed algorithm can also work even with only the information from one base station.**

I. Introduction

Fueled with the emerging cloud technologies (cloud computing, cloud storage, etc), software defined network (SDN) together with network functions virtualization (NFV) is transforming the techonology, and business models in telcommunication industry. Several open source platforms (ONAP, XRAN, etc) have emerged to provide a platform to collect data and share technogies among different vendors and operators. As the wireless network is now evolving into 5G, tremendous data will be shared in these open source platform and thus create a lot of opportunity to provide better service using these data. The shared data can include a wide variety of measurements, such as the service throughput of the mobile device, the serving cell, the signal strength, etc. In [5], a novel localization algorithm has been proposed to estimate the location when measurement reports are made in LTE systems. The paper shows that the medium accuracy of 20m can be achieved for outdoor mobibles. When celluar network evloves from 4G LTE to 5G, a lot of changes have been made to support lower latency and higher throughput. While multiple input multiple output (MIMO) has been extensively used in LTE, the most profound advancement of technology adopted in 5G network is the utilization of a high number of (massive) MIMO antennas especially in the high band. With the use of massive MIMO, transmit power can be effectively beamformed and the mobile will report beam related metric to the base station. Therefore, the approach proposed in [5] for LTE can not be used directly for 5G network. In this paper, we propose a new localization algorithm to estimate the location of mobile devices utilizing the new reported metric introduced in 5G network.

[QUESTION: Does localization in 5G provide improved accuracy compared to localization in LTE. Do we need to say something about this?] [Ans: since it is not easy to direct compare 4G vs. 5G, I added one sentence to say the approach proposed for LTe does not apply directly here.]

As the open source platformsi (such as ONAP, XRAN, etc) are still being developed, it provides an opportunity to request different measurement metrics. The goal of this work is two folds: 1). identify the measurements unique to 5G network for localization. 2). estimate the latitude-longitude of the mobile when the measurement record was generated for 5G network.

In this paper, we design the localization algorithm combining the unique RF fingerprinting in 5G network and probablitistic path-tracking used for robot localization. The unique property of 5G network, such as the beam related information is used in this paper. The focus of this paper is on estimating location for outdoor mobiles. At a high level, our approach has two steps for localizing measurement records from outdoor mobiles:

1) Instead of viewing each NR UE measurement data (NUMD) record in isolation, for each mobile, we *stitch* together NUMD records from that mobile over a "session duration" and model it as a suitable Markovian time series. The problem now reduces to identifying locations (states) of the entire path of the mobile.
2) The above solution method assumes that the probabilities characterizing the underlying Markovian structure can be learned. This is done by performing supervised learning using the unique property **[QUESTION: what is this unique property? beamforming?]** of 5G networks. The training data for supervised learning may come from drive test carried out by network providers once 5G is deployed commein the field. However, given that 5G network deployment is not available in the field during this study, we generate the drive training data and testing data from a 5G simulator.

The details of the above two steps are provided in Section IV.

Our main contribution in this paper is that we have proposed a new localization algorithm appliable to 5G network. Based on our best knowledge, this is the first localization algorithm designed for 5G network.

The rest of the paper is organized as follows. Section **??** provides some background and introduces relevant terminologies. Section III presents the problem setting and states the precise localization problem. Section IV presents the main localization algorithm and Section VI describes the extension of the joint localization and observation model estimation algorithm. We present experimental validation in Section VII and finally we conclude in Section VIII.

II. Relevant 5G Terminologies

Though our techniques could apply to any future cellular system, we use 5G New Radio (NR) terminologies for convenience. The terminologies [2] relevant for our purpose are described below.

UE (user equipment): UE refers to the mobile end-device.

Cell: In 5G networks, a cell refers to coverage footprint of a gNodeB transmitter. Since 5G networks are expected operate in diverse spectrum bands, the cell coverage can vary from as low 100m to 5km. Typically, each cell is expected to cover an azimuth range of 120° using sectorized antennas.

gNodeB (gNB): The gNB is the network element that interfaces with the UE and hosts critical protocol layers like PHY, MAC, and Radio Link Control (RLC) etc. Each gNB could have multiple transmission and reception points (TRP). Typical number number of TRPs is 3 for full coverage around the gNB.

Secondary Synchronization Reference Signal Received Power (SS-RSRP): In 5G networks, UEs make certain measurements of received signal strength on the secondary synchronization (SS) signal for each nearby cell transmitter. SS-RSRP is the total measured time-average received power at UE from the SS signals from a *given cell transmitter*.

beam indices: In 5G-NR, there is the concept of a SS/PBCH block on the downink which comprises of a set of symbols (4) during which beamformed transmissions in a particular direction from gNB could take place. Each SS/PBCH block is identified by an index which the UE uses to report back measurements during a given SS/block. The beam indices we refer to in this paper correspond to the SS/PBCH block index and can have a range from 0 to 63 in standards. We use a grid of beams with upto56 possible directions in the performance evaluation section.

RSSI and RSRQ RSSI (Received Signal Strength Indicator) is the total measured received power at the UE over the entire band of operation from *all cell transmitters.* RSRQ of a given cell transmitter at a UE is RSSI scaled by average RSRP (of that cell) per reference symbol.

We will need a figure to show the connection of 5G network to ONAP/xRAN, etc where the measured data is stored.

ONAP This stands for Open Nework Automation Platform and provides comprehensive platform for real time policy diven orchestration and automa- tion of physcial and virtual network functions that will enable software, network, it and cloud providers and developers to rapidly create new services.

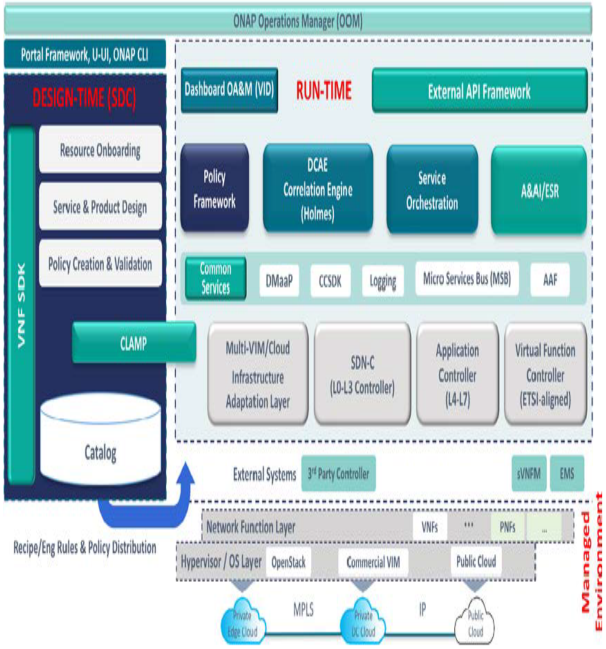


Fig. 1. ONAP architecture

Figure 1 shows the high level architecture. It is expected that in the context of 5G networks and localization topic of this paper, ONAP would orchestrate the localization policy and associated cloud storage within the ”Policy framework”, ”Data Collection, Analytics and Events (DCAE)” and ”Service orchestration” components of the run-time module.

Measurement data collection architecture: The NR UE measurement data (NUMD) collection process occurs primarily in the 5G network via the gNB and Mobility Management Entity (MME). The MME serves as the coordinator of the NUMD data. After NUMD collection is turned on at the gNodeB, it collects the records and sends the data to the MME. MME aggregates and temporarily saves NUMD from multiple gNodeBs and sends it periodically (typically in minutes timescale) to the data center where NUMD is saved and analyzed (for example, ONAP). Scalable storage of NUMD, which can easily run into TB in a week per metropolitan, in the data center is an important design problem and beyond the scope of this paper.

Contents of NUMD: NUMD record contains data related to signaling performance on per UE, per bearer level for different procedures, user experience such as data throughput and procedure duration, gNodeB internal UE related data such as MIMO decision, SINR, buffer size, and normalized power headroom etc. The information present depends on procedure/event that led to the measurement record. For our purpose, we are interested in RF information, and more specifically the SS-RSRP information contained in measurement records.

III. Problem Statement

Consider in an 5G network, mobiles travel along a road network represented by a graph $G_r = (V, E)$ where V denotes graph nodes represented by a latitude-longitude tuple and E denotes valid direct path between two nodes.

There are two two types of data relevant to our discussion:

1) **Training data:** This is essentially geo-tagged data sent from a set of locations in the road graph nodes V. Precisely, we are given n locations $\{x_i\}_{i=1}^n$ and for each location, up to four SS-RSRPs and beam indices of the serving cell. Note that in this study, we assume each UE to report up to four best beam indices and the corresponding SS-RSRPs. We denote by $\{B^j_{i,k}\}_{i=1}^n$, the jth best beam indices sent from training location x_i in cell k and $\{R^j_{i,k}\}_{i=1}^n$ are the corresponding signal strength associated with those reported beam indices. Note that, for a location x_i, the data $R^j_{i,k}$ and $B^j_{i,k}$ are only available for a small subset of cells near location x_i. We will also denote the set of training data by $\mathcal{D}_{tr}$.
2) **NUMD data or observed data:** This data is not geo-tagged but comes with time stamp. Precisely, for every mobile, we are given time instants $t_i, i = 1, 2, \ldots, T$ for each t_i we are also given SS-RSRP $\{\tilde{R}^j_k(t_i)\}_{j=1}^4$ and beam indices $\{\tilde{B}^j_k(t_i)\}_{j=1}^4$ where $k \in K(t_i)$; $K(t_i)$ denotes the set of cells reported by the mobile at time t. Typically $|K(t_i)|$ takes value one or two. Though we have NUMD for each mobile-m, we drop the dependence of m on $R_k(t)$ and $K(t)$ as we are essentially performing the same algorithm for each mobile separately. The locations of mobiles $\tilde{x}(t_i)$ at different times t_i are unknown.

Thus the problem can be stated as follows:

Problem of localization in 5G network: We are given training data consisting of locations $\{x_i\}_{i=1}^n$ and associated SS-RSRPs $\{R^j_{i,k}\}_{i=1}^n$ and beam indices $\{B^j_{i,k}\}_{i=1}^n$of cell-k at location x_i. Assume that the locations are drawn from locations in a road network given by $G_r = (V, E)$. Estimate the unknown location of a sequence of measurements $\tilde{R}^j_k(t_i)$ and $\tilde{B}^j_k(t_i)$ where $i = 1, 2, \ldots, m,\ k \in K(t_i), j = 1, 2, 3, 4$. Note that, in the algorithm illustration in this paper, we only focus on SS-RSRP and beam indices. However, in the field, there may be other measurement reports that can be used to help to improve the localization accuary. For example, timing alignment, etc. These addtional measurement reports can be easily incoorbated into our proposed algorithm.

IV. Localization Algorithms

The framework we use for tackling the localization problem is hidden markov model (HMM) as illustrated in Figure 2. The hidden states in HMM in our case are the locations and the velocity. The observations corresponding to each hidden state are the reported measurement reports, such as SS-RSRP and the beam indices. The system moves from one hidden state to another hidden state with some underlying mobility model. The goal is to infer the hidden state from the observations based on prior knowledge about the transition probabilities between hidden states and observations in the states.

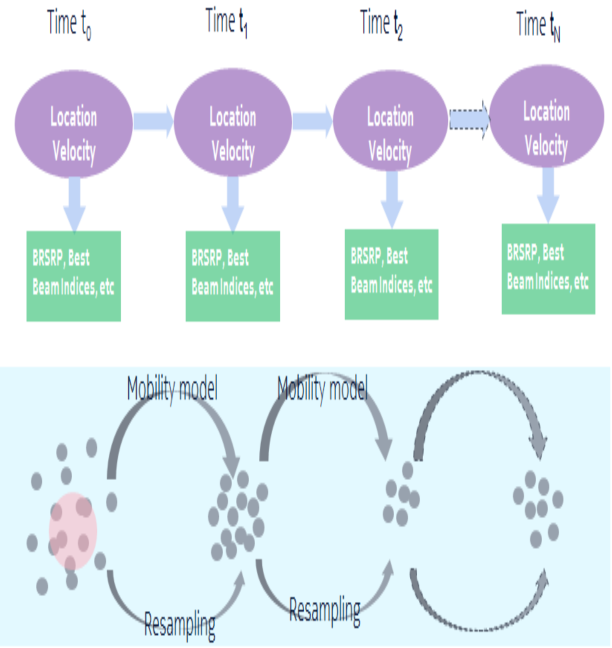


Fig. 2. Hidden Markov Model and Particle filter

In this paper, we use the particle filter to solve the localization problem. The new algorithm $5GLocalizeAlgo$ is stated as follows:

- We first initialize a set of N particles based on some priori distribution, for example, the coverage map of a base station. Here, particles are the samples of the possibles hidden states.
- Each particle has its corresponding weights or likelihoods. We use machine learing algorithms to infer the weights/likelihood as illustrated in Section V.
- Particles move from one state to another state based on the state transition model given in section IV-A.
- The pseudocode is presented in Algorithm 1.

A. State transition probabilities and mobility model

These transition probabilities model how transition happens from one hidden state to another. We assume a suitable mobility model of the mobile which determines how it moves along the graph G_r and also helps us to calculate the above probabilities. Specifically, we assume that the mobile updates its speed according to Gauss-Markov Mobility Model [4].

$$S_t = \alpha S_{t-1} + (1-\alpha)\bar{S} + \sqrt{(1-\alpha^2)}S_{x_{t-1}} \tag{1}$$

$$d_t = \alpha d_{t-1} + (1-\alpha)\bar{d} + \sqrt{(1-\alpha^2)}d_{x_{t-1}} \tag{2}$$

where S_t and d_t are the new speed and direction of the mobile at time interval t. S_{t-1} and d_{t-1} are random variables from a Gaussian distribution with mean $\bar{S}$ and $\bar{d}$.

At each time interval the next location is calculated based on the current location, speed, and direction of movement. Specifically, at time interval t, a mobile's position is given by the equations.

$$x_t = x_{t-1} + S_{t-1} cos(d_{t-1}) * \Delta t \tag{3}$$

$$y_t = y_{t-1} + S_{t-1} sin(d_{t-1}) * \Delta t \tag{4}$$

where (x_t,y_t) and (x_{t-1},y_{t-1}) are the x and y coordinates of the mobile's position at time intervals t and $t+1$, respectively. Δt is the duration of the time interval.

Algorithm 1 $5GLocalizeAlgo(\mathcal{D}_{tr}, G, N_{th})$

```
Offline inference based on training data $\mathcal{D}_{tr}$. This does not need real time processing.
Sample $N$ particles $\mathcal{P}_j = \{\hat{x}_1^{(j)}, \hat{v}_1^{(j)}\}$, $j = 1, \ldots, N$ from prior distribution $p(\tilde{x}_1, \tilde{v}_1 | G)$
Initialize importance weights $\hat{w}_1^{(n)} \leftarrow p(\{\tilde{R}_1^j\}_{j=1}^4, \{\tilde{B}_1^j\}_{j=1}^4 | \hat{x}_1^{(n)})$, $n = 1, \ldots, N$
Normalize $w_1^{(n)} \leftarrow \hat{w}_1^{(n)} / \sum_{l=1}^N \hat{w}_1^{(l)}$, $n = 1, \ldots, N$
for $i = 2$ to $m$ do
    for $n = 1$ to $N$ do
        Sample $\hat{x}_i^{(n)}$ from $\hat{x}_{i-1}^{(n)}$ based on state transition model in section IV-A.
        Update weight $\hat{w}_i^{(n)} \leftarrow \hat{w}_{i-1}^{(n)} \times p(\{\tilde{R}_i^j\}_{j=1}^4, \{\tilde{B}_i^j\}_{j=1}^4 | \hat{x}_i^{(n)})$
    end for
    Normalize $w_i^{(n)} \leftarrow \frac{\hat{w}_i^{(n)}}{\sum_{l=1}^N \hat{w}_i^{(l)}}$
    $\hat{N}_{eff} \leftarrow \frac{1}{\sum_{l=1}^N (w_i^{(l)})^2}$
    if $\hat{N}_{eff} < N_{th}$, $i < m$, resampling condition satisfied then
        Sample $N$ particles with replacement from current particle set $\{\mathcal{P}_j\}_{j=1}^N$ with probabilities $\{\hat{w}_i^{(j)}\}_{j=1}^N$. Update particle set with the new sampled set
        $w_i^{(n)} \leftarrow \frac{1}{N}$ for $n = 1, \ldots, N$
    end if
end for
$n^* = \arg\max_{n=1,\ldots,N} w_m^{(n)}$
Output location estimate $\{\hat{x}_i^{(n^*)}\}_{i=1}^m$
Output distribution
$p(\{\hat{x}^{(n)}\}_{i=1}^m | \{\tilde{R}_i^{1,2,3,4}\}_{i=1}^m, \{\tilde{B}_i^{1,2,3,4}\}_{i=1}^m, G) = w_m^{(j)}$ for $n = 1$ to $N$
```

[QUESTION: in Algorithm 1, what is N_{th}?]

V. Observation modeling using machine learning

As mentioned in the above section, one of the important inference of HMM is the likelihood of observation metric at given state. In this section, we discuss how to infer the likehihood based on the observation metrics, i.e., SS-RSRP and beam indices in a 5G network.

A. Inference on likelihood of beam indices

As part of the observations of HMM model, UE reports up to four observed best beam indices. The probability distribution (also called the likelihood function) of a particular observed beam index such as its j-th ($j = 1, 2, 3, 4$) best reported beam indices at a given location is denoted by $p(\tilde{B}_i^j | \hat{x}_i)$. In our approach, these probabilities can be learnt from the drive test data using machine learning classification algorithms. There are different machine learning classifier in the literature. We use random forest classifier to segment the area into locations where the coverage of beam indices exhibit strong spatial correlation. Note that during the study, we have tried different classifier and random forest indeed presents the best results.

The classification steps are as follows:

1) For each location and each base station that can be heard **[QUESTION: what is heard implying? Sufficient received power?]** at that location, we take the corresponding drive test data beam indices.
2) For each cell, we model the likelihood of a certain classification (beam indices) using *Random Forest* where the latitude and the longitude are taken as features of the model and the likelihood of the beam indices is the output. Each such *Random Forest* is trained using data aggregated in previous step.
3) In this paper, we use the random forest classifier in the machine learning library (sklearn) to predict the probability of reporting a specific beam indices. The function $predict\_proba(X)$ can be used to get the probability of class X.

B. Inference on likelihood of SS-RSRP

The SS-RSRP reported at different states is part of the observations of HMM model. The probability distribution (also called the likelihood function) of an observed SS-RSRP on a location is denoted by $p(\tilde{R}_i | \hat{x}_i)$. In our approach, these probabilities can be learnt from the training data using machine learning regression algorithms. There are different machine learning regression algorithm in the literature. However, SSSPR has the following two special properties:

- The drive test data is spread over a non-contiguous location because coverage areas in a cell are not necessarily connected.
- Wireless SS-RSRP manifests quite different properties in different locations

Thus, we use random forest regressor to segment the area into locations where the SSSPRs exhibit strong spatial correlation.

The regressions steps are as follows:

1) For each location, each base station and each reported set of beam indices that can be heard at that location, we take the empirical mean and standard deviation of all corresponding drive test data SS-RSRP.
2) For each cell and each beam indices, model the spatial variation of SS-RSRP-statistics (i.e., mean and standard

deviation) using *Random Forest* where the latitude and the longitude are taken as features of the model and the SS-RSRP-statistic of the cell is the output. Each such *Random Forest* is trained using data aggregated in previous step. Also, computed is the mean square error (or *cross validation error*) for each random forest.

3) Denote by $RndFrst_m(x, b, c)$ $(RndFrst_s(x, b, c))$ the random forest predictor of mean (standard deviation) of SS-RSRP for cell-c, beam-b at location x. Let $(\sigma_{RF}(b, c))^2$ be the corresponding mean square error of the predictor. Then we model

$$p(\tilde{R}_i^j|\hat{x}_i, \tilde{B}_i^j) = \mathcal{N}(RndFrst(\hat{x}_i, b, c), \sigma_c^2(\hat{x}_i)) \ , \quad (5)$$

where

$$\sigma_c^2(x) = RndFrst_s(x, b, c) + \sigma_{RF}^2(b, c)$$

and the serving cell-c can be obtained from the NUMD record. In general, we can choose any spatial regressor instead of random forest. However, choosing random forest makes the model robust to cell propogation properties and to the fact that the coverage area of the cell could be disjoint.

Note that in LTE network, the regression algorithm is implemented per cell. However, in 5G, with the existence of beam indices, the regression algorithn is now implemented per cell and per beam index.

C. Inference on the likelihood of combined observations

The likelihood of seeing all the observations, $\{\widetilde{R_i^j}\}_{j=1}^4$ and $\{\widetilde{B_i^j}\}_{j=1}^4$ at a given location $\hat{x})i$ is as follows:

$$\begin{aligned} & p(\tilde{R}_i^1, \tilde{R}_i^2, \tilde{R}_i^3, \tilde{R}_i^4, \tilde{B}_i^1, \tilde{B}_i^2, \tilde{B}_i^3, \tilde{B}_i^4|\hat{x}_i) \\ & \quad = \prod_{j=1}^{4} p(\tilde{R}_i^j|\hat{x}_i, \tilde{B}_i^j) p(\tilde{B}_i^j|\hat{x}_i), \end{aligned} \quad (6)$$

VI. Extension: Joint Localization and Channel Modling

In this section we describe our main algorithm for joint observation modelling and UE localization. The Joint Observation Modeling and Localization (JOML) algorithm takes as input the labeled and unlabeled datasets $\mathcal{D}_1, \mathcal{D}_2$, the graph G and a parameter N_{em} specifying the number of expectation-maximization (EM) iterations to be performed. It outputs the UE location estimates $\{\hat{x}_i\}_{i=1}^m$ and the observation model $\mathcal{C}$ in the region of interest. The main idea is to use expectation-maximization procedure to iteratively improve the estimates of both the observation model $\mathcal{C}$ and the location estimates $\{\hat{x}_i\}_{i=1}^m$. The basic EM algorithm improves the channel in each iteration from the previous according to the following equation.

$$\mathcal{C}^{t+1} = \arg\max_{\mathcal{C}} E_{\{\hat{x}_i\}_{i=1}^m|\mathcal{D}_1, \mathcal{D}_2, \mathcal{C}^t} \log P(\mathcal{D}_1, \mathcal{D}_2, \{\hat{x}_i\}_{i=1}^m|\mathcal{C}) \quad (7)$$

Algorithm 2 $JOML(\mathcal{D}_1, \mathcal{D}_2, G, N_{em})$

```
1: $\mathcal{C} \leftarrow ObservationModel(\mathcal{D}_1)$ based on section  V.
2: **for** $j = 1$ to $N_{em}$ **do**
3:     $\{\hat{x}_i\}_{i=1}^m, p(\{\hat{x}_i, \tilde{R}_i^j, \tilde{B}_i^j\}_{i=1}^m|\mathcal{C}, G) \leftarrow$ $5GLocalizeUE(\mathcal{D}_2, \mathcal{C}, G)$ as in Algorithm 1.
4:     $\mathcal{D} \leftarrow \mathcal{D}_1 \cup \{\tilde{R}_i, \hat{x}_i\}_{i=1}^m$
5:     $\mathcal{C} \leftarrow ObservationModel(\mathcal{D})$ based on section  V.
6: **end for**
7: Output $\{\hat{x}_i\}_{i=1}^m = \arg\max_{\{\hat{x}_i\}_{i=1}^m} p(\{\hat{x}_i, \tilde{R}_i, \tilde{R}_i\}_{i=1}^m, |\mathcal{C}, G)$
8: Output $\mathcal{C}$
```

The high level pseudo-code of the algorithm is shown in Algorithm 2.

[QUESTION: I see some inconsistency in the Algorithm 2 description. It is using Algorithm 1. However, Algorithm 1 is taking as input dataset, graph G and $N_th?$. While here it is taking as input dataset, C (a function) and G]

Note that in line 4 $\{\hat{x}_i\}_{i=1}^m$ can be the mode or samples from the distribution $p(\{\hat{x}_i, \tilde{R}_i^j, \tilde{B}_i^j\}_{i=1}^m, |\mathcal{C}, G)$ as a stochastic approximation for the EM algorithm. The JOML algorithm uses two main subroutines. The first subroutine *ObservationModel* computes the likelihood function $\mathcal{C}$ from the labeled dataset $\mathcal{D}$. $\mathcal{C}$ can be viewed as a function which can output the distribution of SS-RSRP and beam indices from any base station k to any UE location $x \in V$. The second subroutine $5GLocalizeUE$ use the observation model function $\mathcal{C}$, the NUMD data $\{\tilde{R}_i, T_i\}$ and the graph G to come up with location estimates of the UE $\{\hat{x}_i\}_{i=1}^m$ and its corresponding distribution as illustrated in section IV. Each time a location estimate is computed it is used with the corresponding record $\tilde{R}_i^j$, $\tilde{B}_i^j$ and the dataset $\mathcal{D}_1$ to further improve the channel model **[QUESTION: why are we calling this a channel model? Are we predicting the channel? Are we not only predicting SS-RSRP and beam indices?]** using an expectation-maximization procedure. Therefore in the next iteration we can obtain a better location estimate.

VII. Evaluation

In this section, we present evaluation of our proposed technique. The objective of our evaluation is three folds **[QUESTION: why are we saying three folds? should it be two fold?]** : to understand the accuracy of our localization scheme, to evaluate how much the accuracy depends of fraction of network coverage area that is drive tested.

A. Methodology

In this paper, we focus on one isolated cell in 5G network, where the measurement report only has the information of the serving cell. In the field, when we have multiple cells, the additional information on the measurement report from neighboring cells can be easily incorporated into the proposed algorithm to improve the localization accuary. This could include information such as RSRQ and RSSI. As there is no commercial deployed 5G network yet, we dont have the

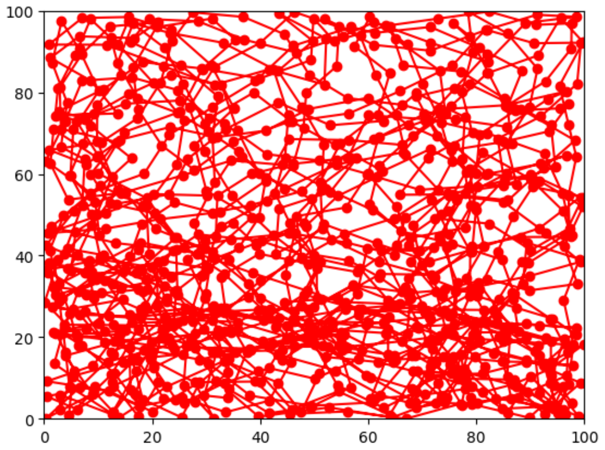


Fig. 3. Training data points

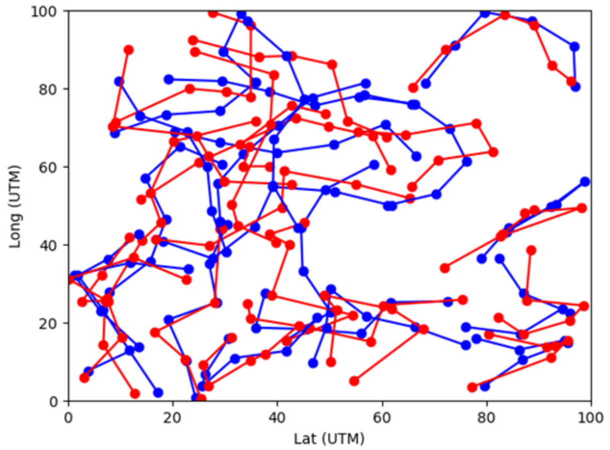


Fig. 4. Comparison of actual (red) and predicted (blue).

drive test data from the field. Instead, we use our 5G system simulator to generate both training data and test data.

The 5G system that we use in this evaluation is operating at mmwave band, and the cell ISD is 100m. The testing data points are illustrated in Figure 3. Instead of diving the training data into training and validation, we use the gridsearch and cross validation functionaility in sklearn [1] to avoid the overfitting since we have limited number of training data.

B. Localization Results

In Figure 4, we show the predicted and actual locations of all mobiles. As it can be seen, the actual locations and the estimated locations are quite close. In the following, we present more detailed analysis of the results.

Inference accuarcy Figure 5(a) and Figure 5(b) illustrate the beam indices at different locations when RF channel conditons are LoS and mixed LoS&NLoS, respectively. Figure 5(c) and Figure 5(d) shows the predicted beam indices using random forest classification. As we can see from these figures, when the training data is more spreaded as shown in Figure 5(b), the random forest classifier will divided the whole area into smaller region. The accuracy score **[QUESTION: should we define what accuracy score means? Is a value of 1 the best?]** for LoS is 0.78 and for mixed LoS&NLoS case is 0.57.

Accuracy CDF: In Figure 6(a) and Figure 6(b), we show the accuracy distribution under two different type of RF channel condition, pure line of sight (LoS) and mixed LoS & non-LoS(NLoS) . More details of these scearions can be found in [3]. With LoS channel, the median accuracy is around $4m$, while with mixed LoS and NLoS channel, the median accuracy is around $12m$.These results have not included additional constraints such as users moving on prescribed paths such as walkways etc. as well as additional columns in the data matrix such as TA, etc. With this additional information, the localization accuracy is expected to improve further.

[QUESTION: should the next para be reworded? it does not seem to connect well to the previous para...]

The rationale behind localizing the path taken by a mobile is two-fold: first, localization accuracy of the individual points can be improved if there is a nearby point that is more accurately localized; and second, we also make use the road network to constrain points to lie on the road whenever the mobile is moving.

VIII. Concluding Remarks

References

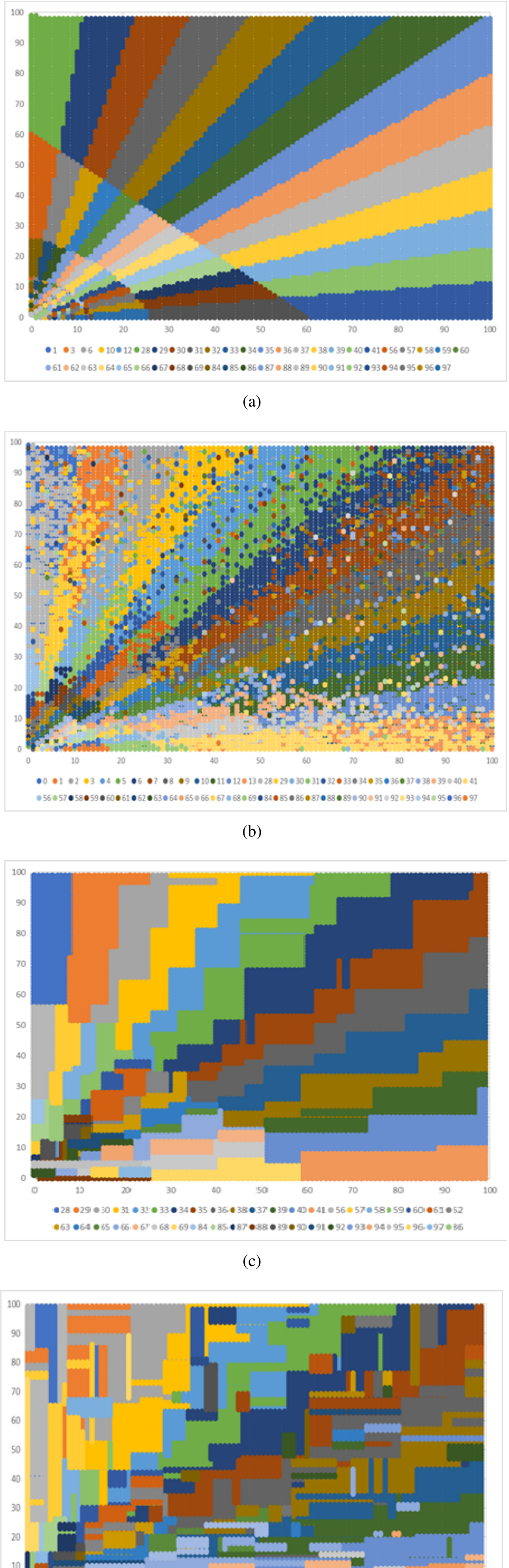
[1] *http://scikit-learn.org/*.

[2] *http://www.3gpp.org/ftp/Specs/2017-12/ Release 15.0.0*.

[3] *http://www.3gpp.org/ftp//Specs/archive/38_series/38.901/38901-e20.zip*.

[4] Camp, T., Boleng, J., and Davies, V. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing, Special Issue: Mobile Ad Hoc Networking – Research, Trends and Applications 2*, 5 (2002), 439—-547.

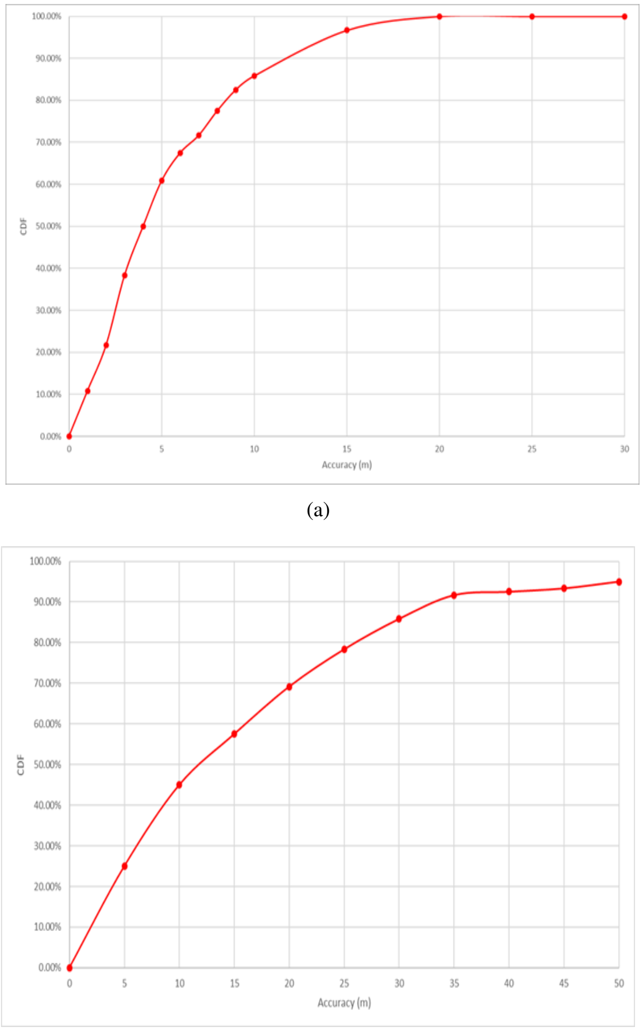
[5] Ray, A., Deb, S., and Mnogioudis, P. Localization of lte measurement records with missing information. In *IEEE INFOCOM 2016* (2016), pp. 1–9.



(a)

(b)

(c)



(a)

(b)

Fig. 6. CDF of accuracy. with different RF channel condition (LoS vs. MixedLoS&NLoS)