LTE Signal Fingerprinting Localization based on CSI

Giovanni Pecoraro, Simone Di Domenico, Ernestina Cianca
Department of Electronics Engineering,
University of Rome Tor Vergata,
Rome, Italy
Email: giovanni.pecoraro@uniroma2.it

Mauro De Sanctis
Department of Electronics Engineering,
University of Rome "Tor Vergata" and
Peoples' Friendship University of Russia "RUDN University"
Email: mauro.de.sanctis@uniroma2.it

***Abstract*—This paper investigates the possibility to use Channel State Information (CSI) extracted from Long Term Evolution (LTE) signals for signal fingerprinting localization. Being the first work in this direction, several types of signal fingerprinting-based approaches have been compared (e.g., CSI-based vs RSSI-based, statistic vs deterministic matching rule). In particular, the paper proposes a novel CSI-based signal fingerprinting that uses as fingerprint not directly the vector of channel gains per subcarrier, but rather some features extracted from these vectors. This method would greatly reduce the memory requirement of the database as well as the computational complexity of the matching phase. Experimental results, shown for both indoor and outdoor environments, confirm the effectiveness of the proposed method and also provide interesting insights on the use of LTE signal fingerprinting based on CSI.**

***Keywords*-localization; fingerprinting; CSI; LTE;**

I. INTRODUCTION

Nowadays, positioning is becoming more and more an essential service. The range of Location-Based Services (LBS) is rapidly expanding and, at the same time, users expect the same level of performance whether they are indoor at home or at work, outdoor in a rural or urban environment, or traveling [1]. The Global Navigation Satellite System (GNSS) represents the most common positioning technology, but it is well known that its accuracy and availability drop in several important application scenarios, such as indoor environments and urban canyons. To enhance positioning accuracy for all types of environment, on one hand there is the effort of the 3rd Generation Partnership Project (3GPP) to support more cellular-based localization techniques, such as Enhanced Cell ID (E-CID), Assisted-Global Navigation Satellite System (AGNSS), Observed Time Difference Of Arrival (OTDOA) and Uplink Time Difference Of Arrival (UTDOA), respectively defined in Releases 9 and 11 of Long Term Evolution (LTE). On the other hand, extensive work has been done in studying alternative localization solutions mainly for indoor environments and based on other radio signals (transmitted by dedicated sensors or by WiFi access points). In both cases, fingerprinting, also known as Radio-Frequency (RF) pattern matching, can play an important role [2]. The basic idea of this technique is to find the location of a mobile device by comparing its signal pattern received from multiple transmitters, such as WiFi Access Points (AP) or cellular Base Stations (BS), to a predefined database of signal patterns. RF pattern matching method is proposed in 3GPP RAN4. Many works have considered a fingerprinting approach for indoor localization using WiFi signals, either based on measurements of the Received Signal Strength Indicator (RSSI) [3] or of the Channel State Information (CSI) [4]. For outdoor environments, most of the proposed signal fingerprint-based localization techniques rely on measurements on the radio signal transmitted by the BSs of a cellular system (GSM [5], WCDMA [6]) or by WiFi APs [7] and the fingerprint is usually a tuple comprising RSSI measurements and BS or AP identifiers (cell-ID or MAC address of the AP). Very recently, also radio measurements from LTE signals have been considered, as in [8] where the performance of a radio fingerprinting localization method that combines LTE and WLAN measurements have been evaluated. The considered radio measurements are the Reference Signal Received Power (RSRP) for LTE and the RSSI for WLAN.

This paper focuses on localization based on CSI measurements of LTE signals. Being, at the best of Authors' knowledge, the first work in this direction (i.e., use of LTE and CSI), the paper aims to answer, through experimental results, to some basic questions such as: ideally, and hence, without considering practical implementation issues related to the costs in terms of time and complexity to build the database of fingerprints, may the use of CSI measurements improve the localization accuracy with respect to signal strength-related metrics (e.g., RSSI)? How to combine the measurements from signals transmitted by different eNodeBs and what is the impact? Moreover, the paper proposes a novel approach in the use of CSI-measurements for fingerprinting, novel with respect to other CSI-based solutions. In most of the previous approaches (mainly proposed for indoor localization and WiFi signals), fingerprints are vectors containing the values of measured CSI. We propose to use as fingerprints not directly the vectors of CSI, but some "descriptors" of the "shape" of the CSI calculated on these vectors. This would greatly reduce the requirements in terms of memory for the database and also the computational complexity of the matching phase. As it will be shown in the paper, the fingerprint could be made up of 2 values per reference point rather than 24. Experimental results, conducted both for indoor and outdoor environments, show the effectiveness of the proposed method, which could be also extended to other types of RF signals (such as WiFi).

The paper is organized as follows: Section II provides the theoretical background on LTE and signal fingerprinting; Section III presents the proposed localization method; Sections IV and V show the experimental setup and results, while the conclusions are drawn in Section VI.

II. Theoretical background

A. LTE Signatures: CSI and RSRP

In case of WiFi, the term CSI indicates the vector of channel gains per subcarrier (usually, 30 subcarriers) which can be also extracted by commodity hardware. In case of LTE, for CSI we still mean a vector of channel gains per subcarriers which represents an estimate of the channel frequency response. Therefore, we first need to explain how to extract these channel gains from the receiver. In the following, we consider only the LTE Frequency Division Duplexing (FDD) mode, in which uplink and downlink channels are separated in frequency. In LTE the information data is spread over a set of orthogonal narrowband subcarriers spaced in frequency by $\Delta f = 15$ kHz. The LTE signal is organized in *frames*, which have a duration of 10 ms. A frame can be represented as a two-dimensional time-frequency grid and consists of 10 subframes. Each subframe is made up of 2 slots of 0.5 ms within which every elementary block, defined as a Resource Element (RE), is mapped to an LTE symbol and subcarrier respectively by its time and frequency indexes. A group of REs on 12 consecutive subcarriers and a slot timeframe is called Resource Block (RB). Let us denote with $\mathbf{y} = \{y_{c,1}, y_{c,2}, ...y_{c,N}\}$ the complex vector, received from one of the 4 possible transmitting antennas of the eNodeB serving the cell c, that, after the N-points FFT at the receiver, corresponds to:

$$\mathbf{y} = \mathbf{X}_c \mathbf{h}_c + \mathbf{w} \tag{1}$$

where $\mathbf{X}_c$ is the transmitted diagonal complex matrix and $\mathbf{h}_c$ is the vector containing the channel complex gains per subcarrier. In Equation 1, noise and inter-cell interference have been modeled as a complex white Gaussian random process $\mathbf{w}$. The channel gains are estimated by the receiver using the Cell Specific Reference Signal (CRS) inserted in some specific Orthogonal Frequency Division Multiplexing (OFDM) symbols within every frame as represented in Figure 1, which shows the CRSs in 1 RB. In Figure 1, CRSs are represented by different colors corresponding to different antennas. Therefore, in 1 slot, there is a total of 4 CRSs per antenna, located over different subcarriers. In this work, we extract the channel gains per subcarrier, using signaling messages within 6 RBs. As we use the signaling messages, it is possible to apply the proposed localization method without having a subscription to a specific operator. Therefore, assuming that the channel stays rather stationary over a slot (0.5 ms), in 1 slot we can extract from the receiver $N = 6*4 = 24$ channel gains for 24 different subcarriers. It is worth noting that at one LTE receiver, it will be possible to extract different vectors $\mathbf{h}_c$, for different transmitting antennas and also for different eNodeBs available in its position.

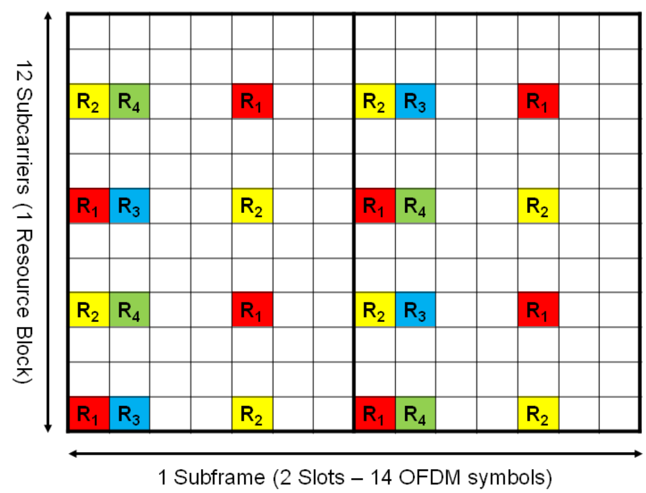

Figure 1. Cell Specific Reference Symbols for each transmitting antenna in 2 consecutive slots and 1 RB

In the fingerprinting approach proposed and investigated in this paper, the used fingerprints will be related to these vectors of channel gains per subcarrier, which represent the CSI. We will also consider the possibility to use as fingerprint the RSRP, which can be calculated on the vectors $\mathbf{h}_c$ as follows:

$$\text{RSRP} = \frac{1}{N} \sum_{i=0}^{N-1} |h_c(i)|^2 \tag{2}$$

B. Signal Fingerprint-based Localization

Signal fingerprint-based localization techniques find the location of a device by comparing its signal pattern received from multiple transmitters (e.g. WiFi APs or cellular BSs) to a predefined database of signal patterns. There is a variety of signal fingerprint-based localization systems in literature both for outdoor and indoor localization. The ”signatures” are usually extracted from the signal received from an AP or a BS of a cellular system. Moreover, in most of the previous works, the fingerprint is a tuple comprising the AP or BS identifiers and the corresponding measured RSSI values. In this work, we propose a signal fingerprinting localization method, which uses as fingerprints specific features calculated on the CSI estimated by an LTE receiver. In general, a fingerprint-based localization consists of the following phases:

- Fingerprint Database Building Phase - The purpose of this phase is to build up the offline fingerprint database, which stores for each Reference Point (RP) a fingerprint. The fingerprint is obtained by properly processing some measurements of the signal received by an AP or a BS. Let us denote with $\mathbf{RF}_r$ the reference fingerprint in the RP r.
- Fingerprint Acquisition Phase - For each Test Point (TP), whose position is unknown, the fingerprint is calculated using the same measurements on the received signal. Let us denote with **TF** the fingerprint in a TP.
- Fingerprint Matching Phase - This phase consists of associating to the fingerprint measured in the TP the fingerprint stored in the built database which is closest to the measured one according to a predefined matching rule. The user location is then calculated as the location of the RP corresponding to the found fingerprint.

There are two main approaches for the matching phase: *Deterministic Matching* and *Statistic Matching*.

Deterministic Matching: For Deterministic Matching, the RP $\overline{r}$, that is associated to the TP whose fingerprint is **TF**, in case of Nearest Neighbor (NN) matching, is the RP that minimizes a deterministic function called **Fingerprint Distance** (FD):

$$\overline{r} : \mathrm{FD}(\mathbf{RF}_{\overline{r}}, \mathbf{TF}) \leq \mathrm{FD}(\mathbf{RF}_r, \mathbf{TF}), \quad \forall r \neq \overline{r} \tag{3}$$

then the location (x, y) of the TP is calculated through the following association:

$$(x, y) \implies (x_{\overline{r}}, y_{\overline{r}}) \tag{4}$$

Other possible deterministic matching methods are the K-Nearest Neighbor (KNN), in which the coordinates of the K best RPs are arithmetically averaged:

$$(x, y) = \frac{1}{K}\sum_{r=1}^{K}(x_r, y_r) \tag{5}$$

and the K-Weighted Nearest Neighbor (KWNN) that computes the weighted average of the coordinates of RPs with the shortest distance to the TP and where the weights correspond to the inverse of the fingerprint distances D_r:

$$(x, y) = \frac{\sum_{r=1}^{K} \frac{1}{D_r}(x_r, y_r)}{\sum_{r=1}^{K} \frac{1}{D_r}}, \quad D_r = \mathrm{FD}(\mathbf{RF}_r, \mathbf{TF}) \tag{6}$$

Statistic Matching: In case of Statistic Matching, it is employed a Naive-Bayes classifier working under the hypothesis of statistical independent features and of equally probable locations. The RP $\overline{r}$ associated to the TP is thus the one maximizing the likelihood function:

$$\overline{r} = \arg\max_r \left[\mathbb{P}\left(r|\mathbf{TF}\right)\right] \tag{7}$$

$$\mathbb{P}\left(r|\mathbf{TF}\right) \propto \prod_{i=0}^{F-1} \mathbb{P}_{f_i}\left(f_i|r\right) \tag{8}$$

Other classifier could be taken under analysis, e.g. Support Vector Machine (SVM), but the Naive-Bayes is preferred because of its simplicity.

III. CSI FINGERPRINTING

In this Section, we explain more in detail the proposed CSI-based approach. As already mentioned, the basic LTE signal measurements extracted from the LTE receiver are the vectors $\mathbf{h}_c$ of $N = 24$ elements as described in Section II. From these vectors, calculated by the same receiver on the LTE signals transmitted by different antennas of multiple eNodeBs, a database like the one shown in Table I is built.

RP	**Coordinates**	**Cell 1**	...	**Cell c**	...	**Cell C**
1	(x_1, y_1)	$\mathbf{R}_{1,1}$	...	$\mathbf{R}_{1,c}$	...	$\mathbf{R}_{1,C}$
⋮	⋮	⋮		⋮		⋮
r	(x_r, y_r)	$\mathbf{R}_{r,1}$	...	$\mathbf{R}_{r,c}$	...	$\mathbf{R}_{r,C}$
⋮	⋮	⋮		⋮		⋮
R	(x_R, y_R)	$\mathbf{R}_{R,1}$	...	$\mathbf{R}_{R,c}$	...	$\mathbf{R}_{R,C}$

Table I
REFERENCE FINGERPRINT DATABASE

In Table I, for each RP r, and for each eNodeB c (denoted as **Cell**), whose signal is received in the RP, a reference fingerprint $\mathbf{R}_{r,c}$ is calculated. We distinguish two main approaches for calculating $\mathbf{R}_{r,c}$:

Direct CSI: $\mathbf{R}_{r,c}$ is directly a vector containing all vectors $\mathbf{h}_{c,t}$, which are the CSIs estimated on the signal received from the cell c and the Tx antenna t. Therefore, considering the c-th eNodeB with $T_c = 4$ antennas, $\mathbf{R}_{r,c}$ is given by:

$$\mathbf{R}_{r,c} = [\mathbf{R}_{r,c,1}, \mathbf{R}_{r,c,2}, \mathbf{R}_{r,c,3}, \mathbf{R}_{r,c,4}] \tag{9}$$

where:

$$\mathbf{R}_{r,c,t} = \mathbf{h}_{c,t} \tag{10}$$

and then the whole reference fingerprint is:

$$\mathbf{RF}_r = [\mathbf{R}_{r,1}, \ldots, \mathbf{R}_{r,c}, \ldots, \mathbf{R}_{r,C}] \tag{11}$$

Obviously, the same procedure is followed for every test fingerprint **TF**. This is how traditionally CSI fingerprinting approaches work. In case of deterministic matching, according to Equation 3, it is also necessary to define a fingerprint distance metric FD, which in case of direct CSI comparison has been chosen equal to:

$$\mathrm{FD}(\mathbf{RF}_r, \mathbf{TF}) = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{T_c}\sum_{t=1}^{T_c} d(\mathbf{RF}_{r,c,t}, \mathbf{TF}_{c,t})\right] \quad (12)$$

where d is the Euclidean Distance between two vectors.

Descriptors: $\mathbf{R}_{r,c}$ is a vector containing M features calculated on the vector $\mathbf{h}_{c,t}$ or on the RSRP. Each feature is a number which is somehow related to the statistics of the Reference Signal Received Power (Table II) or to the "shape" of the Channel Frequency Response (Table III). A similar approach has already been successfully employed in [9] to perform device-free crowd counting and occupancy estimation by WiFi.

Descriptor	Formula
Mean	$\mu = \frac{1}{W}\sum_{s=0}^{W-1}\mathrm{RSRP}[s]$
Standard Deviation	$\sigma = \sqrt{\frac{1}{W-1}\sum_{s=0}^{W-1}(\mathrm{RSRP}[s]-\mu)^2}$
Fano Factor	$\mathrm{FF} = \frac{\sigma^2}{\mu}$

Table II
RSRP DESCRIPTORS

Descriptor	Formula
Mean	$\mu = \frac{1}{N}\sum_{n=0}^{N-1} h_n$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=0}^{N-1}(h_n-\mu)^2}$
Fano Factor	$\mathrm{FF} = \frac{\sigma^2}{\mu}$
Spectral Centroid	$f_n = n15\text{ kHz}, \quad n = [-6\mathrm{RB} : 3 : 6\mathrm{RB}-1]$ $\mathrm{SC} = \frac{\sum_{n=0}^{N-1} h_n f_n}{\sum_{i=0}^{N-1} h_i}$
Spectral Alpha	$\alpha = \sum_{n=0}^{N-1} h_n f_n$
Spectral Lambda	$\lambda = -\frac{1}{N-1}\sum_{n=1}^{N-1}\frac{h_n-h_{n-1}}{f_n-f_{n-1}}\frac{2}{h_n+h_{n-1}}$
Spectral Entropy	$\mathrm{SE} = -\sum_{n=0}^{N-1}\frac{h_n}{\sum_{j=0}^{N-1}h_j}\log_2\frac{h_n}{\sum_{i=0}^{N-1}h_i}$
Spectral Flatness	$\mathrm{SF} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} h_n}}{\frac{1}{N}\sum_{n=0}^{N-1} h_n}$
Spectral Slope	$\mathrm{SSL} = \frac{\sum_{n=0}^{N-1}(f_n-\overline{f}_n)(h_n-\mu)}{\sum_{n=0}^{N-1}(f_n-\overline{f})^2}$
Spectral Spread	$\mathrm{SSP} = \sqrt{\frac{\sum_{n=0}^{N-1} h_n(f_n-\mathrm{SC})^2}{\sum_{i=0}^{N-1} h_i}}$
j-th order Spectral Moment	$\eta_j = \frac{\sum_{n=0}^{N-1} h_n f_n^j}{\sum_{i=0}^{N-1} h_i}$
j-th order Spectral Central Moment	$\xi_j = \frac{\sum_{n=0}^{N-1} h_n(f_n-\mathrm{SC})^j}{\sum_{i=0}^{N-1} h_i}$
Spectral Kurtosis	$T_n = \frac{f_n-SC}{\sqrt{\xi_2}}$ $\mathrm{SKU} = \frac{\sum_{n=0}^{N-1} h_n T_n^4}{\sum_{i=0}^{N-1} h_i}$
Spectral Skewness	$\mathrm{SSK} = \frac{\sum_{n=0}^{N-1} h_n T_n^3}{\sum_{i=0}^{N-1} h_i}$

Table III
CSI DESCRIPTORS

Since descriptors are heterogeneous quantities, which can vary in very different intervals, before performing the deterministic classification it is fundamental to normalize the fingerprints in order to balance all the terms for distance calculation. A *min-max normalization* approach is applied to both reference and test fingerprints:

$$\hat{\mathbf{R}}_{r,c,t} = \frac{\mathbf{R}_{r,c,t} - \min_r[\mathbf{R}_{r,c,t}]}{\max_r[\mathbf{R}_{r,c,t}] - \min_r[\mathbf{R}_{r,c,t}]}, \qquad \forall c,t \quad (13)$$

$$\hat{\mathbf{T}}_{c,t} = \frac{\mathbf{T}_{c,t} - \min_r[\mathbf{R}_{r,c,t}]}{\max_r[\mathbf{R}_{r,c,t}] - \min_r[\mathbf{R}_{r,c,t}]}, \qquad \forall c,t \quad (14)$$

and, in this case, the fingerprint distance is calculated as the vector distance between the normalized fingerprints:

$$\mathrm{FD}(\hat{\mathbf{RF}}_r, \hat{\mathbf{TF}}) = d(\hat{\mathbf{RF}}_r, \hat{\mathbf{TF}}) \quad (15)$$

The statistical matching, instead, is applied only to the approach based on the descriptors since the estimation of the probability distribution functions for all the CSI subcarriers for each antenna relative to all LTE cells is extremely computationally demanding. Moreover, since no distances are calculated by the classifier, the fingerprint normalization is no more needed in such an approach.

The descriptor approach, which is novel, has two fundamental advantages:

- it reduces the amount of data that must be stored in the database. As it will be shown later, with $M = 2$ already very good performance can be achieved. Therefore, for each RP and cell, just 2 elements instead of 24 must be stored;
- it reduces the computational complexity associated to

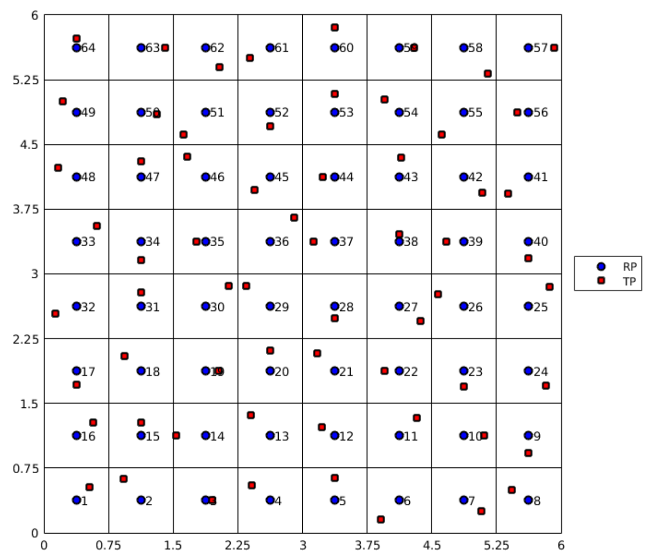

Figure 2. Experimental Outdoor Area of 6 m $\times$ 6 m with a mesh of 0.75 m $\times$ 0.75 m. Blue points represent the Reference Points (RP), while red ones are the Test Points (TP)

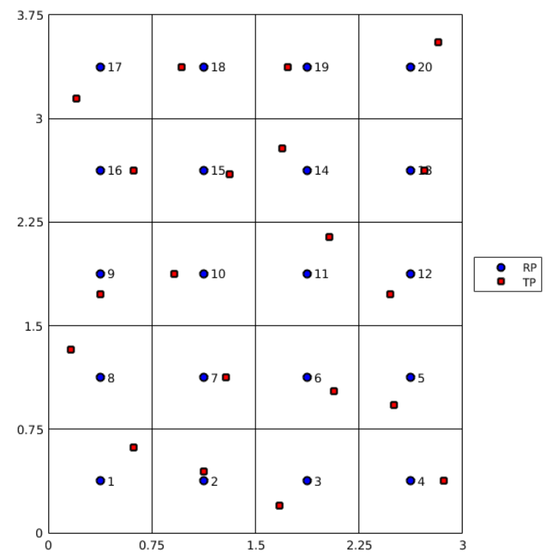

Figure 3. Experimental Indoor Area of 3 m $\times$ 3.75 m with a mesh of 0.75 m $\times$ 0.75 m. Blue points represent the Reference Points (RP), while red ones are the Test Points (TP)

the matching phase, both in case of deterministic and statistic matching.

IV. Experimental Set-up

Localization experiments have been conducted over two different areas:

- an outdoor area of 6 m $\times$ 6 m (36 m^2) in an urban canyon-like environment
- an indoor area of 3 m $\times$ 3.75 m (11.25 m^2) in a small apartment

Both areas have been gridded by a mesh of 0.75 m $\times$ 0.75 m resulting in 64 small squares for the outdoor environment and 20 small squares for the indoor one, as shown respectively in Figure 2 and Figure 3. Reference Points (in blue) have been placed in the center of each square and have been used to build up the offline fingerprint database, while Test Points (in red) have been randomly chosen all over the areas of interest. The number of TPs has been chosen to have a TP for each small square and thus 64 TPs outdoor and 20 TPs indoor. The employed receiver is the Realtek Software Defined Radio RTL2832U (RTL-SDR). The sampling rate and frequency tuning capabilities of the radio platform has allowed to exploit the LTE signal in the 800 MHz Band over a bandwidth of 1 MHz. The choice of the grid size is motivated by the fact that the multipath characteristics, which are caught by the measured CSI, can be considered uncorrelated approximately after one wavelength λ, which is about 37.5 cm. In detail, in both environments the RTL-SDR omni-directional receiving antenna has been placed on the ground in each RP and TP and signal samples relative to the LTE cells 49 (796 MHz), 497 (806 MHz) and 116 (816 MHz) have been captured. While performing the experiments in the outdoor area, it was not possible to completely avoid human presence, in fact there were people moving on the balconies and climbing the stairs of the surrounding building. The indoor experiments, instead, were performed in an apartment without people and where doors and furniture were in fixed positions. The raw samples have then been processed by an LTE software receiver, that, at such sampling rate, has allowed to observe 6 RBs and to get $N = 24$ complex channel gains. Since the just mentioned LTE cells are all configured to work with 2 transmitting antennas, each fingerprint include data from both of them.

V. Experimental Results

For the deterministic methods, a time interval of 1 second, corresponding to $W = 2000$ slots, has been considered to incoherently average the captured CSIs and to calculate RSRP descriptors. In the statistic approach, the same training interval of 1 second has been employed, but in this case it has been selected a smaller sliding windows of just $W = 32$ slots to average the CSI and to calculate the RSRP descriptors. From the LTE data captured on each one among the TPs one has extracted 100 temporally spaced fingerprints resulting in a total of $N_{\mathrm{TF}} = 6400$ outdoor and $N_{\mathrm{TF}} = 2000$ indoor test fingerprints.

In order to verify the effectiveness of the proposed method, the localization performances have also been evaluated in relation to the standard RF fingerprinting localization algorithms, which are essentially based on the RSSI measured values. The RSSI deterministic method consists of generating the reference and the test fingerprints by averaging the RSSI at the considered frequencies, while the probabilistic one is based again on the employment of a Naive-Bayes classifier. Both these methods have been applied in the same experimental conditions and with the same training interval of 1 second. The localization performance index taken under analysis is the **Mean Distance Error (MDE)**. It has been calculated by averaging the Euclidean distance between all the TPs true position $(x_{n,\text{true}}, y_{n,\text{true}})$, which has been set equal to the nearest RP, and the position $(x_{n,\text{est}}, y_{n,\text{est}})$ to which they have been matched:

$$\text{MDE} = \frac{1}{N_{\text{TF}}} \sum_{n=1}^{N_{\text{TF}}} \sqrt{(x_{n,\text{true}} - x_{n,\text{est}})^2 + (y_{n,\text{true}} - y_{n,\text{est}})^2} \tag{16}$$

The MDEs for all the Nearest Neighbor deterministic and probabilistic approaches and for all the combinations of eNodeBs are shown in Table IV for both outdoor and indoor environments. Obviously, the absolute accuracy is strictly related to the choice of the grid size. However, as already stated, the purpose of these results is not to evaluate the absolute accuracy of the method, which will be related also to implementation issues, but rather to compare different types of signal fingerprinting approaches based on LTE signals measurements in order to get better insights into this approach.

A. Outdoor Environment

Results summarized in Table IV show that in an outdoor environment all CSI-based approaches have better performance in terms of localization accuracy with respect to both the standard deterministic and statistic methods based on RSSI. Moreover, deterministic methods show more accurate localization performance than the statistic ones. For all methods, the localization accuracy is strictly dependent on the analyzed LTE cell, but increasing the number of fingerprinted cells can be very helpful. In fact, passing from one LTE signal (received by only 1 eNodeB) to three signals, in case of RSSI deterministic approach, provides an improvement of 16%, while more relevant is the 24%-improvement in case of CSI deterministic approach. In both cases the improvements are calculated with respect to their worst case represented by LTE cell 497. It is interesting to note that the deterministic approach using descriptors has better performance than direct CSI vectors. It is worth noting that such a good performance is achieved by selecting only two descriptors. By employing just the descriptors represented by the **CSI Mean** and the **RSRP Standard Deviation** combined by the NN approach, an accuracy of 1.90 m has been achieved. Therefore, it is possible to work with vectors of only 2 elements rather than 24, allowing a substantial database compression. A further accuracy improvement can also be given by employing the KNN approach, which provides the same positioning accuracy of WKNN, but it is computationally simpler during the matching phase. Figure 4 describes the MDE behavior for all the deterministic methods by varying the number of neighbors. A K increase determines an initial drop and then a floor in MDE, but at the end it causes a degradation in localization accuracy since farther points start to join the weighted average. The descriptor method is always better than both RSSI and CSI methods. By considering $K = 3$ neighbors and employing the same descriptors of NN approach, it is possible to achieve an MDE of 1.68 m, that is 4% more accurate than CSI and 11% more accurate than the best RSSI result.

B. Indoor Environment

Also in case of an indoor environment (see Table IV), CSI-based approaches are better than the RSSI-based ones. Moreover, in this case, using three eNodeB rather than one only provide an improvement of 11% in case of RSSI-based fingerprinting against an improvement of almost 50% in case of CSI-based fingerprinting (compared to their respective worst cases of cells 116 and 49). The NN approach based on the descriptors provides an accuracy of 1.03 m, which is not better with respect to the use of direct CSI vectors. However, still a descriptors-based approach allows to greatly reduce the size of the offline database. Figure 5 describes the MDE behavior for all the KNN deterministic methods by varying K. An increase in the number of neighbors determines a strong improvement in the RSSI deterministic approach, but, at the same time, greatly reduces the localization performance of the CSI method. On the other hand, the descriptor method is much more stable to a K increase and it is able to achieve an MDE of 0.99 m with $K = 5$ and employing again the **CSI Mean** and the **RSRP Standard Deviation** descriptors.

VI. Conclusions

This paper investigates the possibility to use the CSI extracted from LTE signals for signal fingerprinting localization. Experimental results have been conducted in both an indoor and an outdoor environment for different types of approaches. From the comparison results, we can conclude that:

1) CSI-based signal fingerprinting provides better accuracy both indoor and outdoor with respect to RSSI-based approach (more traditionally used for fingerprinting localization).
2) The possibility to combine signals received from different eNodeBs greatly helps to improve the performance. Moreover, it is worth noting that: in case of CSI-based fingerprinting, the improvement in indoor is

	Cells	Fingerprint Type	Classification Mode	MDE	Descriptors
OUTDOOR	497	RSSI	Deterministic - Euclidean	2.94 m	
	497	RSSI	Statistic - Naive Bayes	3.24 m	
	497	CSI	Deterministic - Euclidean	2.77 m	
	497	DESCRIPTORS	Deterministic - Euclidean	2.44 m	CSI Spectral Kurtosis - RSRP Standard Deviation
	497	DESCRIPTORS	Statistic - Naive Bayes	2.47 m	RSRP Standard Deviation - RSRP Fano Factor
	116	RSSI	Deterministic - Euclidean	2.63 m	
	116	RSSI	Statistic - Naive Bayes	3.13 m	
	116	CSI	Deterministic - Euclidean	2.34 m	
	116	DESCRIPTORS	Deterministic - Euclidean	2.11 m	CSI Spectral Lambda - CSI Mean
	116	DESCRIPTORS	Statistic - Naive Bayes	2.40 m	CSI Spectral Kurtosis - RSRP Mean
	49	RSSI	Deterministic - Euclidean	2.64 m	
	49	RSSI	Statistic - Naive Bayes	2.53 m	
	49	CSI	Deterministic - Euclidean	2.64 m	
	49	DESCRIPTORS	Deterministic - Euclidean	2.47 m	CSI Spectral Alpha - RSRP Standard Deviation
	49	DESCRIPTORS	Statistic - Naive Bayes	2.47 m	CSI Mean - RSRP Fano Factor
	497 - 116	RSSI	Deterministic - Euclidean	2.58 m	
	497 - 116	RSSI	Statistic - Naive Bayes	2.89 m	
	497 - 116	CSI	Deterministic - Euclidean	2.42 m	
	497 - 116	DESCRIPTORS	Deterministic - Euclidean	2.04 m	CSI Spectral Lambda - RSRP Mean
	497 - 116	DESCRIPTORS	Statistic - Naive Bayes	2.35 m	CSI Spectral Kurtosis - RSRP Standard Deviation
	497 - 49	RSSI	Deterministic - Euclidean	2.57 m	
	497 - 49	RSSI	Statistic - Naive Bayes	2.64 m	
	497 - 49	CSI	Deterministic - Euclidean	2.54 m	
	497 - 49	DESCRIPTORS	Deterministic - Euclidean	2.11 m	RSRP Mean - RSRP Standard Deviation
	497 - 49	DESCRIPTORS	Statistic - Naive Bayes	2.18 m	RSRP Standard Deviation - RSRP Fano Factor
	116 - 49	RSSI	Deterministic - Euclidean	2.52 m	
	116 - 49	RSSI	Statistic - Naive Bayes	2.47 m	
	116 - 49	CSI	Deterministic - Euclidean	2.08 m	
	116 - 49	DESCRIPTORS	Deterministic - Euclidean	2.00 m	CSI Mean - RSRP Fano Factor
	116 - 49	DESCRIPTORS	Statistic - Naive Bayes	2.11 m	RSRP Mean - RSRP Fano Factor
	497 - 116 - 49	RSSI	Deterministic - Euclidean	2.47 m	
	497 - 116 - 49	RSSI	Statistic - Naive Bayes	2.53 m	
	497 - 116 - 49	CSI	Deterministic - Euclidean	2.11 m	
	497 - 116 - 49	DESCRIPTORS	Deterministic - Euclidean	1.90 m	CSI Mean - RSRP Standard Deviation
	497 - 116 - 49	DESCRIPTORS	Statistic - Naive Bayes	2.24 m	RSRP Standard Deviation - RSRP Fano Factor
INDOOR	497	RSSI	Deterministic - Euclidean	1.51 m	
	497	RSSI	Statistic - Naive Bayes	1.45 m	
	497	CSI	Deterministic - Euclidean	1.58 m	
	497	DESCRIPTORS	Deterministic - Euclidean	1.08 m	4th-order CSI Spectral Moment - RSRP Standard Deviation
	497	DESCRIPTORS	Statistic - Naive Bayes	1.22 m	RSRP Standard Deviation - RSRP Fano Factor
	116	RSSI	Deterministic - Euclidean	1.34 m	
	116	RSSI	Statistic - Naive Bayes	1.39 m	
	116	CSI	Deterministic - Euclidean	1.95 m	
	116	DESCRIPTORS	Deterministic - Euclidean	1.26 m	CSI Spectral Entropy - 2nd CSI Spectral Moment
	116	DESCRIPTORS	Statistic - Naive Bayes	1.26 m	4th CSI Spectral Central Moment - CSI Fano Factor
	49	RSSI	Deterministic - Euclidean	1.64 m	
	49	RSSI	Statistic - Naive Bayes	1.69 m	
	49	CSI	Deterministic - Euclidean	1.35 m	
	49	DESCRIPTORS	Deterministic - Euclidean	1.21 m	4th CSI Spectral Central Moment - RSRP Standard Deviation
	49	DESCRIPTORS	Statistic - Naive Bayes	1.43 m	RSRP Mean - RSRP Standard Deviation
	497 - 116	RSSI	Deterministic - Euclidean	1.29 m	
	497 - 116	RSSI	Statistic - Naive Bayes	1.29 m	
	497 - 116	CSI	Deterministic - Euclidean	1.27 m	
	497 - 116	DESCRIPTORS	Deterministic - Euclidean	1.01 m	RSRP Mean - RSRP Fano Factor
	497 - 116	DESCRIPTORS	Statistic - Naive Bayes	1.29 m	RSRP Standard Deviation - RSRP Fano Factor
	497 - 49	RSSI	Deterministic - Euclidean	1.56 m	
	497 - 49	RSSI	Statistic - Naive Bayes	1.33 m	
	497 - 49	CSI	Deterministic - Euclidean	1.22 m	
	497 - 49	DESCRIPTORS	Deterministic - Euclidean	1.00 m	RSRP Standard Deviation - RSRP Fano Factor
	497 - 49	DESCRIPTORS	Statistic - Naive Bayes	1.18 m	4th CSI Spectral Moment - 4th CSI Spectral Central Moment
	116 - 49	RSSI	Deterministic - Euclidean	1.45 m	
	116 - 49	RSSI	Statistic - Naive Bayes	1.51 m	
	116 - 49	CSI	Deterministic - Euclidean	1.54 m	
	116 - 49	DESCRIPTORS	Deterministic - Euclidean	1.20 m	CSI Spectral Flatness - RSRP Standard Deviation
	116 - 49	DESCRIPTORS	Statistic - Naive Bayes	1.46 m	4th CSI Spectral Central Moment - RSRP Standard Deviation
	497 - 116 - 49	RSSI	Deterministic - Euclidean	1.46 m	
	497 - 116 - 49	RSSI	Statistic - Naive Bayes	1.34 m	
	497 - 116 - 49	CSI	Deterministic - Euclidean	0.97 m	
	497 - 116 - 49	DESCRIPTORS	Deterministic - Euclidean	1.03 m	CSI Spectral Kurtosis - RSRP Standard Deviation
	497 - 116 - 49	DESCRIPTORS	Statistic - Naive Bayes	1.30 m	3th CSI Spectral Central Moment - 4th CSI Spectral Central Moment

Table IV
MDE OF ALL NN DETERMINISTIC AND PROBABILISTIC APPROACHES FOR ALL ENODEB COMBINATIONS

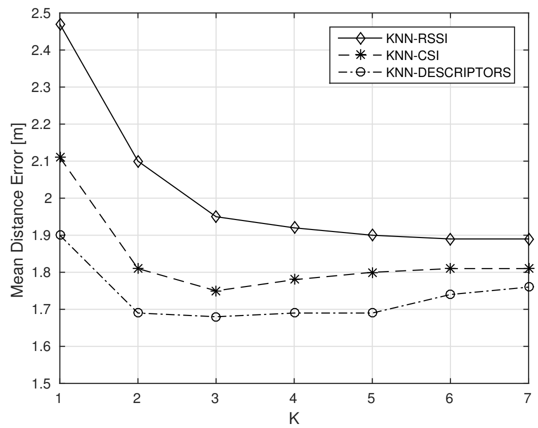

Figure 4. Outdoor environment - MDE of KNN deterministic approaches for different K values

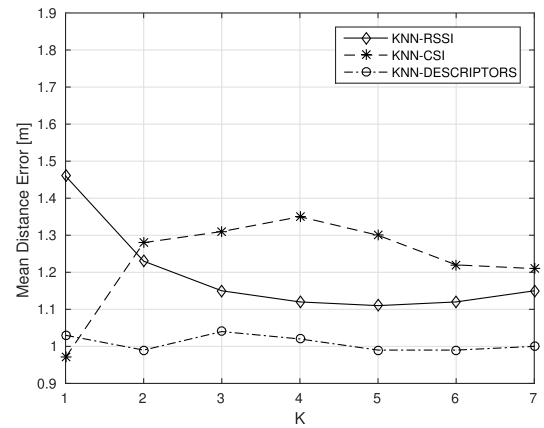

Figure 5. Indoor environment - MDE of KNN deterministic approaches for different K values

higher than outdoor (50% w.r.t. 24%, respectively); in case of RSSI-based fingerprinting, diversity provided by different eNodeB does not provide much improvement in an indoor environment.

3) The use of deterministic approaches, which are also less computationally complex than statistical ones, provide better localization performance.
4) Furthermore, the results achieved by the proposed method based on the use of CSI descriptors rather than CSI vectors are very promising. By employing only two descriptors, i.e. CSI Mean and RSRP Standard Deviation, and by properly choosing the value of K in the KNN approach, it is possible to achieve at least the same accuracy as the more traditional method based on CSI vectors. However, the offline database can be compressed by a factor of 12 in both outdoor and indoor environments.
5) It is also worth noting that the CSI Mean is linked to the shape of the CSI and takes into account the specific multipath environment around the reference point. The RSRP Standard Deviation, instead, is still calculated from the CSI but, as a matter of fact, it is strictly linked to the signal strength and thus to the distance from the eNodeB.

Finally, we can conclude that using CSI from LTE signals could be worth both for indoor and outdoor localization. Nevertheless, the feasibility of the approach is also strictly related to implementation issues (which might be critical in an outdoor scenario) that will be investigated in future works.

Acknowledgment

The publication was partially financially supported by the Ministry of Education and Science of the Russian Federation (the Agreement number 02.a03.21.0008).

References

[1] Ericsson, ”Positioning with LTE”, September 2011, 284 23-3155 Uen

[2] Vo, Quoc Duy, and Pradipta De. ”A survey of fingerprint-based outdoor localization”, *IEEE Communications Surveys & Tutorials* 18.1 (2016): 491-506.

[3] P. Bahl and V. Padmanabhan, RADAR: an in-building rf-based user location and tracking system in INFOCOM 2000. *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2000, pp. 775784 vol.2

[4] Wu, Kaishun, Jiang Xiao, Youwen Yi, Dihu Chen, Xiaonan Luo and Lionel M. Ni. CSI-Based Indoor Localization. IEEE Transactions on Parallel and Distributed Systems 24 (2013): 1300-1309.

[5] M. Ibrahim and M. Youssef, Cellsense: An accurate energy-efficient gsm positioning system, *Vehicular Technology, IEEE Transactions on*,vol. 61, no. 1, pp. 286296, Jan 2012.

[6] B. Jakub, and L. Jukka, ”Pilot correlation positioning method for urban UMTS networks,” in *Proc. European Wireless Conference*, vol. 2, pp. 464-469,APR.2005.

[7] J. Rekimoto, T. Miyaki, and T. Ishizawa, Lifetag: Wifi-based continuous location logging for life pattern analysis in *Location- and Context-Awareness*, ser. Lecture Notes in Computer Science, J. Hightower, B. Schiele, and T. Strang, Eds., vol. 4718. Springer Berlin Heidelberg, 2007, pp. 3549.

[8] J. Turkka, T. Hiltunen, R. U. Mondal and T. Ristaniemi, ”Performance evaluation of LTE radio fingerprinting using field measurements”, *Wireless Communication Systems (ISWCS), 2015 International Symposium on.* IEEE, 2015

[9] S. Di Domenico, G. Pecoraro, E. Cianca, M. De Sanctis, ”Trained-once device-free crowd counting and occupancy estimation using WiFi: A Doppler spectrum based approach”, *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, vol. 00, no. , pp. 1-8, 2016