7

Low-Complexity Offline and Online Strategies for Wi-Fi Fingerprinting Indoor Positioning Systems

Giuseppe Caso, Luca De Nardis, Maria-Gabriella Di Benedetto

DEPARTMENT OF INFORMATION ENGINEERING, ELECTRONICS AND TELECOMMUNICATIONS (DIET), SAPIENZA UNIVERSITY OF ROME, ROME, ITALY

1 Introduction

Indoor localization of wireless mobile devices, also referred to as indoor positioning, is nowadays an intensively investigated research topic, toward the extension of outdoor location-based services to indoor environments. Among the available communication technologies and infrastructures, Wi-Fi appears as an excellent localization support, since it is largely widespread in indoor environments, implying low implementation time and costs (Liu et al., 2007).

Fingerprinting is one of the most investigated techniques for the implementation of Wi-Fi indoor positioning systems (IPSs) (Honkavirta et al., 2009). It relies on a preliminary collection of location-dependent signal propagation data at predefined positions in the area of interest, called reference points (RPs). Received signal strength (RSS) collection is common in implementing Wi-Fi fingerprinting IPSs, and leads to defining the *fingerprint* as the set of RSS values, measured in a given RP, from nearby Wi-Fi access points (APs) (Bahl and Padmanabhan, 2000).

RSS fingerprinting is typically organized in two phases: offline and online. In the offline phase, an RSS fingerprint is collected at each RP, in order to create a discrete radiomap of the area. In the online phase, the unknown position of a target device is estimated as a function of the position of RPs, that best matches the *online reading*, that is the RSS fingerprint measured by the target device.

Accuracy and complexity of fingerprinting mainly depend on two factors: (1) a careful planning of the offline phase, particularly in terms of RP locations and number of measurements, and (2) an optimized definition of the estimation algorithm used in the online phase.

As regards the offline phase, the goal is to achieve a satisfactory trade-off between positioning accuracy, and efforts and time dedicated to the collection of RPs. Previous work highlighted the impact of human body and device orientation on measured RSS values, and the need for multiple measurements at each RP, in order to counteract channel variability and measurement errors (Hossain et al., 2007; Liao and Kao, 2008; Honkavirta et al., 2009; Kessel and Werner, 2011). Furthermore, in order to limit the measurement effort, two main approaches have been proposed:

- *RSS prediction*: Most of the RSS values are predicted, rather than measured. RSS prediction may be used either for the generation of *virtual* RPs, leading to *discrete virtual* fingerprinting, or for the evaluation of continuous RSS distributions, leading to *continuous virtual* fingerprinting. RSS prediction may be obtained by either *indoor propagation modeling*, in which an empirical radio propagation model, trained with a set of initial measurements, is used, or by *interpolation*, in which adjacent real RPs are interpolated (Chintalapudi et al., 2010; Kumar et al., 2016; Hernández et al., 2017).
- *Crowdsourcing*: System users also contribute to the collection of RSS fingerprints (Bolliger, 2008). Application of crowdsourcing entails a further challenge, that is heterogeneity, in terms of RP locations and devices used for collection, as they both depend on users (Laoudias et al., 2013).

Considering the online phase, the goal is to achieve a satisfactory trade-off between positioning accuracy and algorithm complexity, by decreasing the average number of online operations required for position estimation. As regards the online procedure, deterministic (Bahl and Padmanabhan, 2000; Shin et al., 2012; Caso et al., 2015b), and probabilistic (Roos et al., 2002; Youssef and Agrawala, 2004; Le Dortz et al., 2012) weighted k-nearest neighbors (WkNN) schemes are, by far, the most widely investigated. In particular, deterministic WkNN is highly appealing, since it requires the evaluation of a simple deterministic similarity metric between the online reading and each RP fingerprint. Previous work highlighted the impact on the achievable accuracy of the value of k, and the similarity metric used for RPs selection and weighting (Caso et al., 2015b). As regards the reduction of the online operations, the adoption of *two-step* algorithms was proposed (Youssef et al., 2003; Feng et al., 2012; Yu et al., 2014), in which:

- The offline phase is organized in RPs *collection* and *clustering*, during which measurements are collected and then divided into nonoverlapping groups, according to the adopted similarity criterion.
- The online phase is organized in *coarse* and *fine* localization. During coarse localization, the online reading is compared with a fingerprint associated with each cluster, according to the adopted similarity metric; only the RPs within clusters passing a predefined similarity threshold are used in fine localization, where a traditional WkNN estimator can be then applied.

The main goal of two-step algorithms is to reduce the online complexity, by reducing the RP space. While for a generic WkNN *flat* algorithm, in fact, all RPs are compared with

the online reading, in order to select the k most relevant ones, in two-step algorithms, only RPs belonging to selected clusters are taken into account. Among others, the use of Affinity Propagation (Frey and Dueck, 2007), that evaluates RPs mutual similarities, has been proposed for the creation of clusters (Feng et al., 2012; Tian et al., 2013). As a result, in Affinity Propagation two-step algorithms, the definition of the similarity metric has an important role at RP clustering, and coarse and fine localization phases.

This chapter analyzes and discusses the trade-off between accuracy and complexity, within offline and online phases of Wi-Fi fingerprinting IPSs. The main goal is to identify and derive low-complexity strategies, leading to a simple system implementation, while preserving the achievable positioning accuracy.

As regards the offline phase, a strategy adopting RSS prediction, in the form of discrete virtual fingerprinting via indoor propagation modeling, is proposed, and compared with traditional, real RPs only, deterministic WkNN (Caso and De Nardis, 2015, 2017; Caso et al., 2016). As for online phase, Affinity Propagation two-step deterministic WkNN is compared with traditional flat deterministic WkNN, in order to highlight factors, that mainly impact on system complexity and accuracy (Caso et al., 2015a).

Experimental results reported in the present work were obtained in the testbed implemented at the first two floors of the Department of Information Engineering, Electronics, and Telecommunications (DIET) of Sapienza University of Rome.

The chapter is organized as follows: Sections 2 and 3 present the proposed low-complexity strategies for offline and online phases, respectively. In both sections, the reference model is first introduced, followed by a description of the testbed, experimental settings, and performance indicators, and a discussion on the obtained results. Section 4 highlights main results, and the advantage of using the proposed low-complexity strategies, and concludes the chapter underlying possible future research lines.

2 Low-Complexity Strategy for Offline Phase

This section focuses on the description of the proposed low-complexity strategy for the offline phase. As introduced in Section 1, a discrete virtual fingerprinting approach via indoor propagation modeling is implemented, in order to decrease the effort in RSS measurements, while maintaining a satisfying positioning accuracy. In particular, the empirical multiwall multifloor (MWMF) indoor propagation model is adopted for the creation of virtual RPs, while a traditional deterministic WkNN estimator is used in the online phase.

2.1 RSS Prediction via MWMF Model

The MWMF model (Damosso, 1999) is an appealing solution for indoor propagation modeling, due to the good trade-off between simplicity and prediction accuracy (Borrelli et al., 2004). MWMF takes into account objects that may obstruct the propagation over an indoor link, leading to the following path loss model (Damosso, 1999):

Table 1 A_{MWMF} Parameters' Description

Parameters	Description
l_c	Constant loss
N_{obj}	Number of different families of 2D objects
I_n	Number of types of 2D objects considered for family n
$N_{n,i}$	Number of 2D obstructing objects of family n and type i
N_f	Number of obstructing floors
$l_{n,i}$	Loss due to 2D objects of family n and type i
l_f	Loss due to obstructing floors
b	Empirical 3D propagation parameter

$$\mathrm{PL}_{\mathrm{MWMF}} = \mathrm{PL}_{\mathrm{OS}} + A_{\mathrm{MWMF}} \quad (\mathrm{dB}), \tag{1}$$

where $\mathrm{PL}_{\mathrm{OS}}$ models the path loss over the transmitter-receiver (Tx-Rx) distance d, while A_{MWMF} models the additional loss due to obstructing obstacles, that may be different in nature and type, such as walls, doors, pipelines, and others. Details on both terms of Eq. (1), with a particular focus on the peculiar MWMF term, that is A_{MWMF}, are provided in Caso et al. (2016). As regards A_{MWMF}, a linear combination of different obstructing obstacles is considered (Borrelli et al., 2004), and a generic description of parameters within this term is reported in Table 1.

The use of the MWMF model requires an initial, hopefully small, set of M measurements, and also, for each mth measurement ($m = 1, 2, \ldots, M$), the information regarding number, type, and positions of objects obstructing the Tx-Rx direct path, indicated as the set of topological parameters $\{\mathcal{T}_m\}$. The measurements are used in order to estimate the set of propagation parameters $\{\mathcal{S}\}$, within $\mathrm{PL}_{\mathrm{OS}}$ and A_{MWMF} terms, that characterizes the model in the area of interest, such as path loss exponent and loss terms due to obstructing objects. An iterative least square fitting procedure can be adopted, so to minimize the difference between RSS initial measurements and predictions in the same M positions, and in turn estimate $\{\mathcal{S}\}_{\mathrm{opt}}$, as follows:

$$\{\mathcal{S}\}_{\mathrm{opt}} = \underset{\{\mathcal{S}\}}{\mathrm{argmin}} \left\{ \sum_{m=1}^{M} |\mathrm{RSS}_m - \widehat{\mathrm{RSS}}_m|^2 \right\}, \tag{2}$$

where, for the mth available measurement, RSS_m and $\widehat{\mathrm{RSS}}_m$ are the actual versus predicted RSS values at the Rx, when considering a Tx emitting a known effective isotropic radiated power (EIRP) $\mathrm{W}_{\mathrm{Tx}}^{\mathrm{EIRP}}$ at distance d_m. At each iteration, $\widehat{\mathrm{RSS}}_m$ is computed as the difference between $\mathrm{W}_{\mathrm{Tx}}^{\mathrm{EIRP}}$ and the path loss evaluated as in Eq. (1), and the iterative procedure stops when the difference with RSS_m is minimized, and thus $\{\mathcal{S}\}_{\mathrm{opt}}$ is found.

The propagation parameters to be optimized, included in $\{\mathcal{S}\}$, may differ depending on the model definition, and may include parameters from both $\mathrm{PL}_{\mathrm{OS}}$ and A_{MWMF} terms; the set adopted in this work is reported in Caso et al. (2016).

2.2 Offline Phase

Given a set of L Wi-Fi APs at known positions in the area of interest $\mathcal{A}$, initial measurements in a set of N^{r} real RPs are collected, so that an $L \times 1$ RSS fingerprint $\boldsymbol{s}_{n_1}$ is associated with the n_1th RP ($n_1 = 1, 2, \ldots, N^{\text{r}}$). The generic $\boldsymbol{s}_{n_1}$ component, denoted by s_{l,n_1}, contains the RSS measured by a reference Rx, placed at the n_1th RP, from the lth AP. This value is usually obtained by averaging $q > 1$ repeated measurements, in order to counteract propagation channel variability and measurement errors.

The MWMF model is then used for the generation of virtual RPs. The model is first calibrated on the set of N^{r} real RPs, so that the derived propagation parameters are used for the generation of further N^{v} virtual RPs. The component $\hat{s}_{l,n_2}$ of the generic $L \times 1$ fingerprint $\hat{\boldsymbol{s}}_{n_2}$ contains the predicted RSS at the n_2th virtual RP from the lth AP ($n_2 = 1, 2, \ldots, N^{\text{v}}$).

Both the amount and the positions of real RPs are expected to affect the generation of virtual RPs and the achievable positioning accuracy, since the amount, in particular, defines the number of measurements considered in the model fitting procedure of Eq. (2). Given N^{r} real RPs, regularly spaced over a grid in the environment $\mathcal{A}$ of area $|\mathcal{A}|$, their spatial density is given by:

$$d^{\text{r}} = \frac{N^{\text{r}}}{|\mathcal{A}|}, \tag{3}$$

that defines the number of real RPs for each meter square of $\mathcal{A}$.

In Caso and De Nardis (2015, 2017) and Caso et al. (2016), several strategies were proposed for the use of real RPs in Eq. (2), toward the evaluation of the propagation parameters for the MWMF model; in this work, a *Specific AP Fitting* strategy is adopted, where the measurements at all RPs are discriminated considering the AP they refer to, so that a different set of propagation parameters is obtained for each AP.

In analogy to Eq. (3), denoting with N^{v} the number of virtual RPs to be generated, their spatial density is given by:

$$d^{\text{v}} = \frac{N^{\text{v}}}{|\mathcal{A}|}. \tag{4}$$

Position of virtual RPs in $\mathcal{A}$ can be freely defined: in this work a grid placement is adopted, as an extension of the commonly adopted real RPs placement.

2.3 Online Phase

A deterministic WkNN estimation algorithm, using the combination of real versus virtual RPs, is used to infer the target location. Denoting by $N = N^{\text{r}} + N^{\text{v}}$ the total number of RPs in $\mathcal{A}$, $\boldsymbol{s}_n$ the RSS fingerprint of nth RP ($n = 1, 2, \ldots, N$), and $\boldsymbol{s}_i$ the RSS online reading collected during the ith positioning request by a target device in unknown position $\boldsymbol{p}_i = (x_i, y_i, z_i)$, the position estimate relies on the computation of a set of similarity metrics $\text{sim}_{n,i} = \text{sim}(\boldsymbol{s}_n, \boldsymbol{s}_i)$. The W$k$NN algorithm selects the k RPs that present the highest $\text{sim}_{n,i}$ values and provides an estimate of $\boldsymbol{p}_i$ defined as:

$$\hat{p}_i = \frac{\sum_{n=1}^{k}(\text{sim}_{n,i})p_n}{\sum_{n=1}^{k} \text{sim}_{n,i}}, \tag{5}$$

where $p_n = (x_n, y_n, z_n)$ is the position of the nth selected RP, and $\hat{p}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$ is the estimated position of the target device. The similarity metric can be any deterministic metric computable in the RSS space between vectors s_n and s_i. A popular choice is the inverse Minkowski distance of order o, defined as follows:

$$\text{sim}_{n,i} = [\mathcal{D}_{n,i}^o]^{-1} = \left[\left(\sum_{l=1}^{L}|s_{l,i} - s_{l,n}|^o\right)^{1/o}\right]^{-1}, \quad o \geq 1. \tag{6}$$

Typical orders are $o = 1$ (inverse Manhattan distance) and $o = 2$ (inverse Euclidean distance). Similarity metrics based on modified versions of the inner product between RSS vectors have also been proposed (Torres-Sospedra et al., 2015; Caso et al., 2015a). In this work, the inverse Euclidean distance is adopted as similarity metric.

As regards the value of k, it is selected as follows:

$$k = \lceil 0.05(d^r + d^v)|\mathcal{A}|\rceil, \tag{7}$$

that is proved to be a reliable estimator of the optimal value of k, defined as the one minimizing the average positioning error (Caso et al., 2016).

2.4 Experimental Setting and Performance Indicators

Experimental analysis of the proposed low-complexity strategy was conducted in the testbed at the DIET of Sapienza University of Rome. The testbed is deployed in an office environment, and covers two floors with an area of approximately $42 \times 12 \text{ m}^2$ each. $L_1 = 6$ Wi-Fi APs working @2.4 and 5 GHz and $L_2 = 7$ Wi-Fi APs working @2.4 GHz, with a beacon transmission period of $T_b = 100$ ms and a transmit power $W_{Tx}^{EIRP} = 20$ dBm, are placed at known positions at the first and second floors, respectively. On both floors, APs are placed in the false ceiling of the central corridor, due to deployment constraint.

The second floor was adopted as area of interest $\mathcal{A}$, and the APs on this floor were considered in the fingerprinting measurement campaign, so that $L = L_2$. During the offline phase, $N^{r,tot} = 72$ RPs were selected for RSS collection on a regular grid within $\mathcal{A}$; fingerprints were also collected in a set of $N^t = 31$ target points (TPs) randomly distributed over $\mathcal{A}$. TPs were used as ground truth, in order to test the positioning accuracy of the proposed scheme. In both cases, fingerprints were obtained as the average of $q = 50$ scans at each location. Furthermore, all measurements were carried out during weekend afternoons, using a MacBook Pro equipped with an AirPort Extreme Network Interface Card, placed on a wooden platform.

Two different analyses were carried out in order to demonstrate the advantage of using the proposed low-complexity strategy:

- *Analysis I*: Reliability of the MWMF model in generating virtual RPs.
- *Analysis II*: Effect of adopting the proposed virtual fingerprinting scheme on the achievable positioning accuracy.

As regards Analysis I, MWMF reliability was evaluated as a function of the parameter ρ, that determines the number N^r of RPs used for the model fitting of Eq. (2), out of the total number of measured RPs, $N^{r,tot}$, so that $N^r = \lceil \rho N^{r,tot} \rceil$ (Caso et al., 2016). For each considered value of ρ, the MWMF model was fitted, and RSS values in the $N^{r,tot}$ RP locations were generated and compared with the collected ones, so to evaluate the prediction error, as follows:

$$\delta_{l,n}(\rho) = |s_{l,n} - \hat{s}_{l,n}(\rho)|, \tag{8}$$

where $\hat{s}_{l,n}(\rho)$ is the predicted RSS for the generic (AP$_l$, RP$_n$) pair, obtained by using a set of $N^r = \lceil \rho N^{r,tot} \rceil$ RPs in the model fitting procedure, while $s_{l,n}$ is the measured RSS for the same pair. Assuming the generic $\delta_{l,n}(\rho)$ value as a sample of a random variable $\delta_l(\rho)$ related to the lth AP, the cumulative distribution function (CDF) of $\delta_l(\rho)$, that is $F_{\delta_l(\rho)}(\delta_{l,n}(\rho)) = \Pr\{\delta_l(\rho) \leq \delta_{l,n}(\rho)\}$, and the average error $\bar{\delta}_l(\rho) = \frac{\sum_{n=1}^{N^{r,tot}} \delta_{l,n}(\rho)}{N^{r,tot}}$ were also evaluated.

Considering Analysis II, positioning accuracy of the proposed strategy was evaluated as a function of densities d^r and d^v, and adopting a value of k as in Eq. (7). The analysis was carried out by computing the positioning error $\epsilon_i(d^r, d^v)$ for each TP i ($i = 1, 2, \ldots, N^t$) as follows:

$$\epsilon_i(d^r, d^v) = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}, \tag{9}$$

where $(x_i, y_i) = \boldsymbol{p}_i$ and $(\hat{x}_i, \hat{y}_i) = \hat{\boldsymbol{p}}_i$ are the actual and the estimated positions of the ith target device at the DIET second floor, respectively. As in the case of the prediction error $\delta_l(\rho)$, the CDF of positioning error $\epsilon(d^r, d^v)$, that is $F_{\epsilon(d^r, d^v)}(\epsilon_i(d^r, d^v)) = \Pr\{\epsilon(d^r, d^v) \leq \epsilon_i(d^r, d^v)\}$, and the average positioning error $\bar{\epsilon}(d^r, d^v) = \frac{\sum_{i=1}^{N^t} \epsilon_i(d^r, d^v)}{N^t}$ were evaluated.

2.5 Results and Discussions

Fig. 1A shows the CDF of the prediction error $\delta_l(\rho)$ for the AP fitting strategy, a reference AP, and four different values of ρ (note that $\rho = \{0.1, 0.2, 0.5, 1\}$ lead to $N^r = \{8, 15, 35, 72\}$ uniformly distributed RPs used in the model fitting procedure). Results show that slightly different $\delta_l(\rho)$ errors are obtained as ρ increases from 0.1 to 1, suggesting that a relatively small value of ρ (and in turn of N^r and d^r) is sufficient to obtain a reliable estimation of the propagation parameters for the generation of virtual RPs, although a few measurements are still required, due to the empirical nature of the MWMF model. Fig. 1B shows the CDF of $\delta_l(\rho)$ for the AP fitting strategy ($\rho = 1$) against a baseline *No Fit* strategy with no fitting, that is no use of measurements, and by obtaining RSS predictions using propagation parameters estimated for a different area, and reported in Borrelli et al. (2004). Results show that the prediction error significantly increases when no site-specific model fitting is carried out.
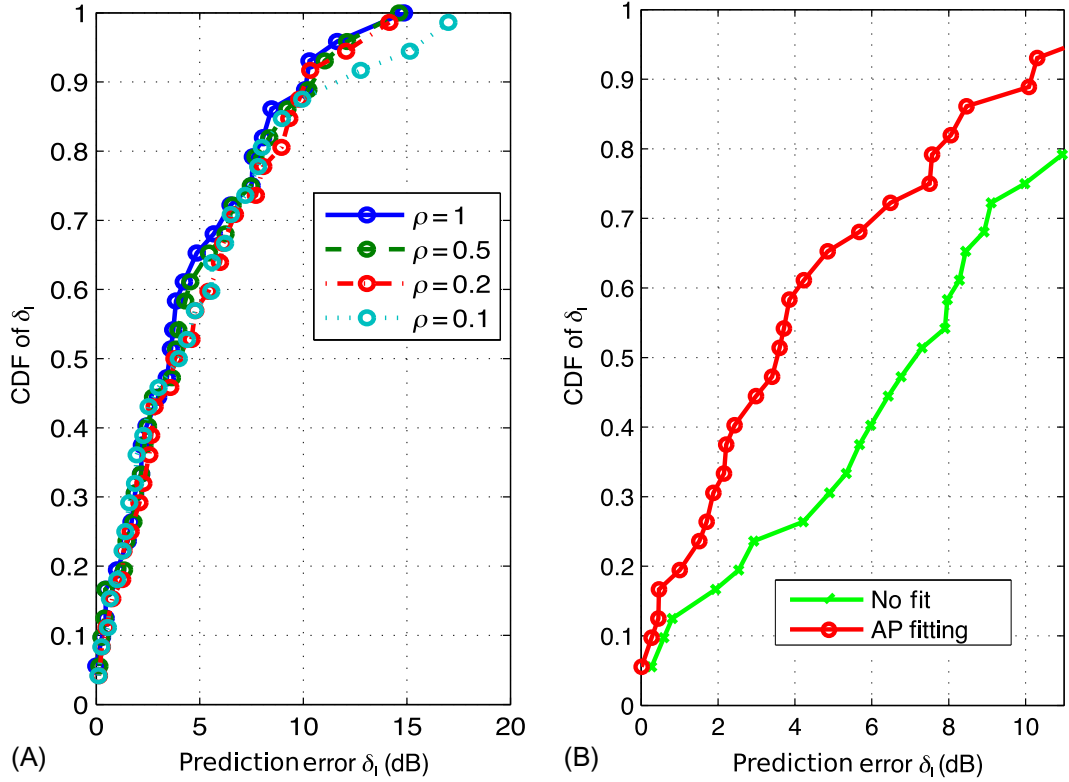
FIG. 1 CDFs of the prediction error $\delta_l(\rho)$ for a reference AP: AP fitting strategy (real RPs selection parameter $\rho = \{0.1, 0.2, 0.5, 1\}$) (A), no fit versus AP fitting strategy ($\rho = 1$) (B).

Fig. 2 shows the average positioning error $\bar{\epsilon}(d^r, d^v)$ of virtual fingerprinting, as a function of d^v, and four different values of d^r, corresponding to the values of ρ and N^r adopted in the previous analysis (see Eq. 3). The value of k is fixed as in Eq. (7). Fig. 2 also reports the average error obtained by adopting real RPs only fingerprinting, with a number of real RPs corresponding to minimum and maximum values of d^r. Results highlight a few important aspects: on one hand, the introduction of virtual MWMF fingerprints does not lead to performance improvement when a large enough amount of real ones is collected (see, in particular, $d^r = 0.14$ as a function of d^v curve vs. $d^r = 0.14$ with $d^v = 0$ one). On the other hand, a significant accuracy improvement is obtained, when a large amount of virtual RPs ($d^v = 10$) was predicted from a limited set of initial real measurements in the offline phase, and used in the online phase (see $d^r < 0.14$ as a function of d^v curves vs. $d^r = 0.02$ and $d^r = 0.14$ with $d^v = 0$ ones). Furthermore, note that the error obtained with virtual fingerprinting almost accommodates on the error of the optimal real RPs only scheme, where a large amount of real RPs is used. The accuracy decrease is of about 60 cm in the extreme case ($d^r = 0.02$ as a function of d^v).
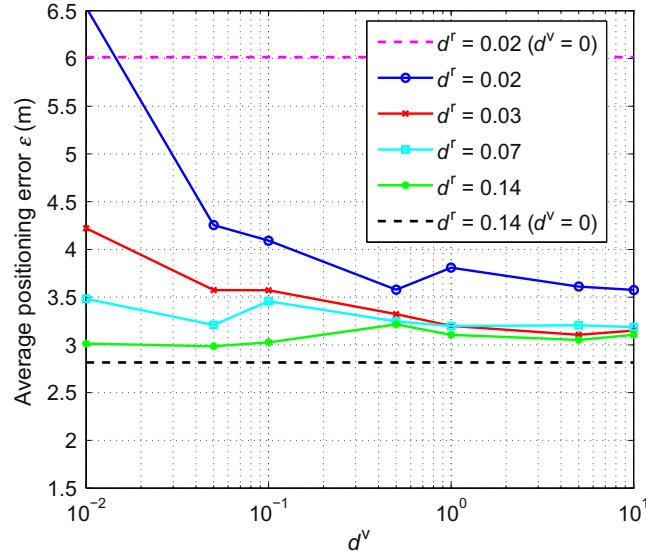
FIG. 2 Average positioning error $\bar{\epsilon}(d^r, d^v)$ for virtual fingerprinting, as a function of virtual RPs density d^v, and four different values of real RPs density d^r; upper and lower bounds of the error are also shown, for real RPs only fingerprinting with $d^r = 0.02$ and $d^r = 0.14$.

Results clearly show that creation and adoption of a large enough amount of virtual RPs can significantly reduce the offline phase complexity, in terms of number of needed initial measurements, while preserving achievable accuracy.

3 Low-Complexity Strategy for Online Phase

This section focuses on the description of the proposed low-complexity strategy for the online phase. As introduced in Section 1, an Affinity Propagation two-step algorithm is implemented, in order to decrease the number of online operations required for a position estimate, while maintaining a satisfying positioning accuracy. In particular, the offline phase foresees an RP clustering step via Affinity Propagation, while the online phase is divided into coarse (cluster selection) and fine (RPs selection and weighting) localization.

3.1 RP Clustering via Affinity Propagation

Affinity Propagation is a clustering algorithm, that divides a set of elements in clusters, and elects for each cluster a representative clusterhead, also dubbed as exemplar (Frey and Dueck, 2007). The algorithm usually follows a distributed and iterative approach: elements are seen as network nodes which exchange messages containing computed values, that measure the *affinity* of one element with another element, until it converges to a stable set of exemplars and corresponding clusters.

In the context of indoor positioning, since elements correspond to RPs, the algorithm is centralized and, for each iteration, the requested values are evaluated by a central processing unit based on an initial measure of similarity $\text{sim}^{\text{CL}}_{n_1,n_2} = \text{sim}^{\text{CL}}(s_{n_1}, s_{n_2})$ (with $n_1, n_2 = 1, 2, \ldots, N$, $n_1 \neq n_2$) between each RP pair, where the superscript CL indicates that such similarity values are evaluated during the clustering step, measuring how well the RP_{n_2} is suited to be the exemplar for RP_{n_1}. The self-similarity value $\text{sim}^{\text{CL}}(s_n, s_n)$ (with $n = 1, 2, \ldots, N$), that is also dubbed as *preference*, indicates the possibility that RP_n may become an exemplar. In order to give all RPs the same chance to become an exemplar, their preferences are initially set to a common finite value, typically defined as:

$$\text{pref}(s_n) = \text{sim}^{\text{CL}}(s_n, s_n) = \gamma \cdot \text{median}\{\text{sim}^{\text{CL}}(s_{n_1}, s_{n_2})\}, \quad \forall n_1, n_2 \in \{1, 2, \ldots, N\}, \quad n_1 \neq n_2,$$

$$(10)$$

where γ is a tunable parameter (Feng et al., 2012; Frey and Dueck, 2007), equal to 1 in the present work.

The definition of exemplars relies on the iterative evaluation of two values between each RP pair:

- *Responsibility* $\text{resp}(s_{n_1}, s_{n_2})$: It reflects the accumulated evidence for how well-suited RP_{n_2} is to serve as the exemplar for RP_{n_1}, taking into account other potential exemplars for RP_{n_1}.
- *Availability* $\text{avail}(s_{n_1}, s_{n_2})$: It reflects the accumulated evidence for how appropriate it would be for RP_{n_1} to choose RP_{n_2} as its exemplar, taking into account the support from other RPs that RP_{n_2} should be an exemplar.

These values are updated according to the following equations:

$$\text{resp}(s_{n_1}, s_{n_2}) = \text{sim}^{\text{CL}}(s_{n_1}, s_{n_2}) - \max_{n_3} \left\{ \text{avail}(s_{n_1}, s_{n_3}) + \text{sim}^{\text{CL}}(s_{n_1}, s_{n_3}) \right\}, \tag{11}$$

$$\text{avail}(s_{n_1}, s_{n_2}) = \min \left\{ 0, \text{resp}(s_{n_2}, s_{n_2}) + \sum_{n_3} \max\{0, \text{resp}(s_{n_3}, s_{n_2})\} \right\}, \tag{12}$$

$\forall n_1, n_2, n_3 \in \{1, 2, \ldots, N\}$, $n_1 \neq n_2$, $n_3 \neq n_2$ in Eq. (11), $n_3 \neq n_1, n_2$ in Eq. (12).

In order to facilitate convergence of the iterative procedure and avoid ringing oscillations, a *damping* factor $\text{DF} \in [0.5, 1)$ is typically introduced leading to the following expressions for the new values of responsibility and availability:

$$\text{resp}_{\text{new}}(s_{n_1}, s_{n_2}) = \text{DF} \cdot \text{resp}_{\text{old}}(s_{n_1}, s_{n_2}) + (1 - \text{DF}) \cdot \text{resp}'_{\text{new}}(s_{n_1}, s_{n_2}),$$
$$\text{avail}_{\text{new}}(s_{n_1}, s_{n_2}) = \text{DF} \cdot \text{avail}_{\text{old}}(s_{n_1}, s_{n_2}) + (1 - \text{DF}) \cdot \text{avail}'_{\text{new}}(s_{n_1}, s_{n_2}),$$

$$(13)$$

$\forall n_1, n_2 \in \{1, 2, \ldots, N\}$, $n_1 \neq n_2$, with $\text{resp}'_{\text{new}}(s_{n_1}, s_{n_2})$ and $\text{avail}'_{\text{new}}(s_{n_1}, s_{n_2})$ evaluated by using Eqs. (11), (12), respectively. $\text{DF} = 0.6$ is generally adopted, and thus also used in the present work.

Two main issues are identified in the application of Affinity Propagation:

- *Degeneracies*: Degeneracies can arise if, for example, the similarity metric is commutative and two elements (RPs) are isolated from all the others. In this case

oscillations in deciding which of the two elements should be the exemplar might appear. The solution proposed in Frey and Dueck (2007), and also adopted in the present work, is to add a small amount of random noise to similarities values to avoid such deadlock situations.

- *Outliers*: When applied to RP clustering, the algorithm might occasionally lead to an RP belonging to a cluster, but being physically far away from the cluster exemplar. In Feng et al. (2012), taking advantage of the knowledge of each RP position, each outlier is forced to join the cluster characterized by the exemplar at minimum distance from the outlier itself. This solution is also adopted in the present work.

3.2 Offline Phase

Affinity Propagation is used for grouping the RPs collected in the offline phase. Given a set of L Wi-Fi APs that can be detected in $\mathcal{A}$ (differently from the assumption given in Section 2.2, position of APs may be unknown and it is not required for the application of the present low-complexity online phase strategy), initial measurements in a set of N RPs are collected, so that an $L \times 1$ RSS fingerprint s_n is associated with the nth RP ($n = 1, 2, \ldots, N$).

After the RSS collection, Affinity Propagation clustering takes place, and the RPs are divided into $N_c < N$ clusters. The definition of similarity, used during the iterative process, may be inherited from Frey and Dueck (2007); in this case, given a pair of RPs, $\text{sim}^{\text{CL}}(s_{n_1}, s_{n_2})$ is as follows:

$$\text{sim}^{\text{CL}}(s_{n_1}, s_{n_2}) = -[\mathcal{D}^2(s_{n_1}, s_{n_2})]^2 \quad \forall n_1, n_2 \in \{1, 2, \ldots, N\}, \quad n_1 \neq n_2, \tag{14}$$

where $\mathcal{D}^2(s_{n_1}, s_{n_2})$ expresses the Euclidean distance between two RP fingerprints. In order to use the Affinity Propagation algorithm in its traditional settings, this definition is also adopted in this work. Detailed analysis on the impact of using different definitions can be found in Caso et al. (2015a).

3.3 Online Phase

Once the offline phase is complete, the position estimate is obtained through coarse and fine localization steps. In coarse localization, $N_{c,i} \leq N_c$ clusters that best match the s_i online reading are selected, through the computation of N_c similarity values $\text{sim}^{\text{C}}_{n_c,i} = \text{sim}^{\text{C}}(s_{n_c}, s_i)$ (with $n_c = 1, 2, \ldots, N_c$) between the online reading and a fingerprint selected as the n_cth cluster's representative, denoted as s_{n_c}. In the present work, s_{n_c} is a synthetic fingerprint, generated by averaging the fingerprints of the RPs within a cluster. The selection of the clusters is performed by comparing each $\text{sim}^{\text{C}}_{n_c,i}$ with a threshold α, defined as follows (Feng et al., 2012):

$$\alpha = \alpha_1 \cdot \max_{e \in \mathcal{E}} \left\{ \text{sim}^{\text{C}}(s_i, e) \right\} + \alpha_2 \cdot \min_{e \in \mathcal{E}} \left\{ \text{sim}^{\text{C}}(s_i, e) \right\}. \tag{15}$$

Clusters with similarity values above α are selected. In Eq. (15), $\mathcal{E}$ denotes the set of cluster fingerprints, and $\alpha_1 + \alpha_2 = 1$. The values of α_1 and α_2 allow to adjust the number of selected

clusters: as an example, the smaller the number of desired selected clusters, the higher should be the value of α_1 (and conversely, the lower the value of α_2). $\alpha_1 = 0.95$ and $\alpha_2 = 0.05$ are the values adopted in the present work.

Denoting as N_i the total number of RPs within the selected $N_{c,i}$ clusters, k out of N_i RPs are then selected during the fine localization, by computing N_i similarity values $\text{sim}^F_{n,i} = \text{sim}^F(s_{n_i}, s_i)$ (with $n_i = 1, 2, \ldots, N_i$) between the online reading and the RP fingerprints. Similarly to Section 2.3, final position estimate is then obtained via WkNN, so that:

$$\hat{p}_i = \frac{\sum_{n=1}^{k} \left(\text{sim}^F_{n,i} \right) p_n}{\sum_{n=1}^{k} \text{sim}^F_{n,i}}. \tag{16}$$

It is worth noting that $\text{sim}^C_{n_c,i}$ and $\text{sim}^F_{n,i}$ are properly defined similarity metrics, and superscripts C and F indicate that such similarities are evaluated during the coarse and fine localization, respectively. Similarly to the discussion in Section 2.3, several definitions may be adopted. In this work, the inverse Euclidean distance is considered at both steps, so that, overall, the proposed Affinity Propagation two-step algorithm adopts the similarity definition of Eq. (14) for RP clustering ($\text{sim}^{CL}_{n_1,n_2}$), while the one of Eq. (6) for coarse and fine localization ($\text{sim}^C_{n_c,i}$ and $\text{sim}^F_{n,i}$), respectively.

Detailed analysis on the impact of using different metric combinations at different steps can be found in Caso et al. (2015a).

As regards parameter k in the fine localization WkNN, a *dynamic k* selection scheme is used, in which the value of k is adjusted at each positioning request, as also proposed in Shin et al. (2012), Marcus et al. (2013), and Caso et al. (2015b). In general, this scheme relies on the definition of a threshold λ taking values in the same domain of the similarity metric, and on the selection of the RPs that show a value of the metric above the threshold. In this case, given the ith positioning request, λ is evaluated as a function of the average on the RPs similarity values $\text{sim}^F_{i,n}$, that is:

$$\lambda_i \left(\text{sim}^F_{n,i} \right) = c \cdot \overline{\text{sim}}^F_{n,i} = c \cdot \frac{\sum_{n=1}^{N} \text{sim}^F_{n,i}}{N}, \tag{17}$$

where c is a tuning parameter, ranging from 0.1 to 2 in the present work.

3.4 Experimental Setting and Performance Indicators

Experimental analysis of the proposed low-complexity strategy was conducted in the testbed implemented at the DIET of Sapienza University of Rome, described in Section 2.4.

In this case $N_1 = 65$ and $N_2 = 69$ RPs were collected, in a grid fashion, at the first and second floors, respectively, for a total number of $N = 134$ RPs. For each RP, RSS values received from all detected APs, were collected, by averaging $q = 5$ measurements. Once the offline stage was completed, the total number of detected APs was $L = 133$, including physical and virtual APs as well as temporary and mobile connection points. In particular L also contained $L_1 = 6$ and $L_2 = 7$ testbed-dedicated Wi-Fi APs at DIET first and second floors, as described in Section 2.4. Fingerprints were also collected in a set of $N^t = 70$ TPs,

randomly distributed at both floors. TPs were used as ground truth, in order to test the positioning accuracy of the proposed scheme. All measurements were carried out during weekdays, using an Android Samsung Tablet held on by a human surveyor.

Two different analyses were carried out in order to demonstrate the advantage of using the proposed low-complexity strategy:

- *Analysis I*: Effect of adopting the proposed two-step scheme on the achievable positioning accuracy.
- *Analysis II*: Effect of adopting the proposed two-step scheme on the complexity of the online phase, in terms of average number of operations required for obtaining a position estimate.

As regards Analysis I, positioning accuracy of two-step versus flat algorithms was evaluated as a function of the threshold parameter c defined in Eq. (17). The flat algorithm corresponds to a traditional WkNN scheme, with no RP clustering, and thus no coarse localization. Only fine localization is used, by comparing the online fingerprint with each RP fingerprint. The analysis was carried out by computing the 3D positioning error $\epsilon_i(c)$ for each TP i ($i = 1, 2, \ldots, N^t$), as follows:

$$\epsilon_i(c) = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2}, \tag{18}$$

where $(x_i, y_i, z_i) = \boldsymbol{p}_i$ and $(\hat{x}_i, \hat{y}_i, \hat{z}_i) = \hat{\boldsymbol{p}}_i$ are the actual and the estimated position of the ith target device in the coordinate system including DIET first and second floors, respectively. CDF of positioning error $\epsilon(c)$, that is $F_{\epsilon(c)}(\epsilon_i(c)) = \Pr\{\epsilon(c) \leq \epsilon_i(c)\}$, and the average positioning error $\bar{\epsilon}(c) = \frac{\sum_{i=1}^{N^t} \epsilon_i(c)}{N^t}$ were also evaluated as a function of c.

Regarding the computational complexity, the selected performance indicator was the number of similarity values N_{sim} to be computed for obtaining a position estimate. In the case of the two-step algorithm, N_{sim} for the generic ith online reading can be expressed as follows:

$$N_{\text{sim}} = N_c + N_i, \tag{19}$$

where N_c is the number of RP clusters and N_i is the number of RPs passing the coarse localization. Noting that in case of the flat algorithm $N_{\text{sim}} = N$ for each positioning request, one can observe that, on average, the adoption of the two-step algorithm will lead to a reduction of computational complexity if $\overline{N}_{\text{sim}} = N_c + \overline{N}_i < N$, where $\overline{N}_{\text{sim}}$ is the average number of similarity computations, depending in turn on the average number of selected RPs $\overline{N}_i = \frac{\sum_{i=1}^{N^t} N_i}{N^t}$.

3.5 Results and Discussion

Before discussing the results of the analysis described in Section 3.4, a brief mention to the results of the RP clustering is reported. By adopting Affinity Propagation on 100 iterations, with $\text{sim}^{\text{CL}}(\boldsymbol{s}_{n_1}, \boldsymbol{s}_{n_2})$ as in Eq. (14), and tuning parameters as reported in Section 3.1, a total number of 13 clusters were obtained in the area of interest, 6 on the first floor, and 7 on

the second one, respectively. Fig. 3 reports the clusters obtained at DIET first floor, with corresponding exemplars. An average of 10 RPs is included in each cluster, with maximum cluster amount of 15 and minimum of 5 RPs.

Two-step and flat algorithms were then compared in terms of positioning accuracy. Fig. 4 shows the average positioning error $\bar{\epsilon}(c)$ for both schemes, as a function of the threshold parameter c, adopted in fine localization WkNN (in case of two-step, cluster selection in the coarse localization was performed by using the threshold α of Eq. 15). Results show that the two-step algorithm always leads to comparable or better results than
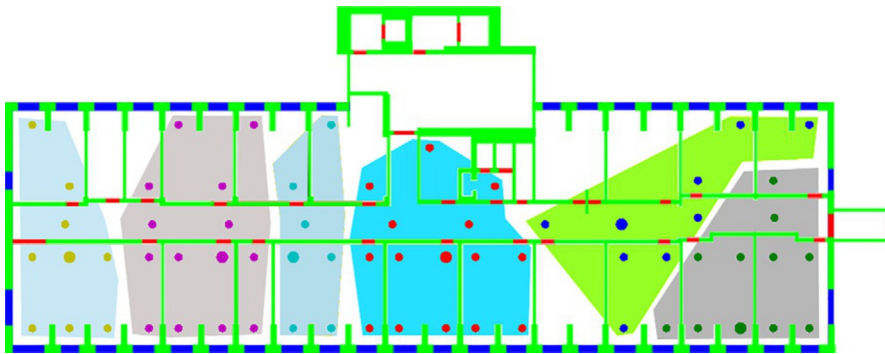

FIG. 3 RP clustering at DIET first floor (areas represented in *different colors* indicate different clusters; *larger dots* indicate exemplars).
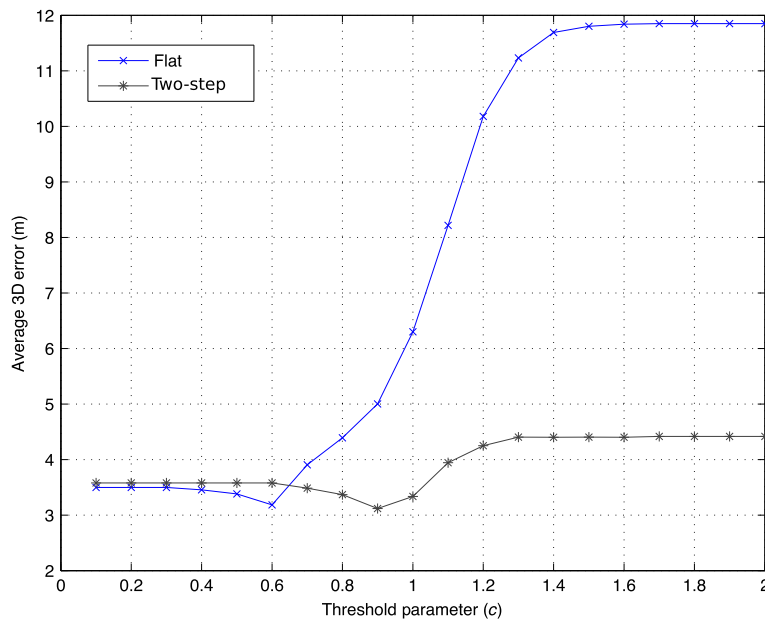

FIG. 4 Flat vs. two-step: 3D average positioning error $\bar{\epsilon}$ as a function of the threshold parameter c.
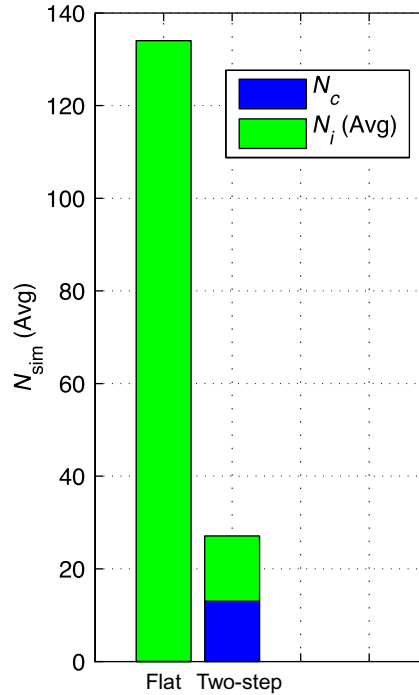
FIG. 5 Flat vs. two-step: Average number of similarity values $\overline{N}_{sim}$ to be computed for obtaining a position estimate.

the flat algorithm. Performance improvement is particularly significant when the adopted value of c leads to thresholds, that allow a large number of RPs to be selected. Under these conditions, the RP space reduction provided by two-step schemes leads to significant reduction in $\overline{\epsilon}$, and in turn improved positioning accuracy.

Fig. 5 shows the average number of similarity values N_{sim} to be computed for obtaining a position estimate, and confirms the main expected advantage of the use of two-step algorithms, that is the reduction of online complexity. While, in fact, the number of computed similarity values is always equal to N for the flat algorithm, a significantly lower average value, of about 27, is obtained for the two-step scheme.

4 Conclusion and Future Work

In this work, a theoretical and experimental analysis of Wi-Fi RSS fingerprinting IPSs was presented, focusing on the trade-off between performance and complexity of offline and online fingerprinting phases. Two low-complexity system implementation strategies were proposed and analyzed in the experimental testbed at the DIET of Sapienza University of Rome.

Considering the offline phase, in order to limit the efforts related to the RSS collection, the use of the MWMF indoor propagation model was proposed for the generation of

virtual fingerprints. The MWMF accuracy in predicting RSS values in the area of interest and the impact of using such values in the online phase were analyzed. Experimental results showed that a significant reduction of measurement collection is possible, since positioning accuracy is preserved thanks to the use of virtual RPs.

As regards the online phase, an Affinity Propagation two-step WkNN algorithm was analyzed. Experimental results showed a significant decrease of required operations for obtaining a position estimate, with a preserved or even improved positioning accuracy, when the two-step algorithm was compared with a traditional flat scheme.

Moving from this work, several research lines can be identified. The joint application of the proposed solutions is under investigation, and is being tested in different environments, in order to generalize the obtained results. Moreover, the analysis may be extended to different positioning approaches, such as continuous virtual fingerprinting in the offline phase, and probabilistic estimation in the online one. Finally, considering recent research trends, the definition of low-complexity strategies for hybrid IPSs, that exploit different wireless technologies in order to provide accurate localization, is expected be extremely important.

References

Bahl, P, Padmanabhan, V.N., 2000. RADAR: an in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 775–784.

Bolliger, P, 2008. Redpin-adaptive, zero-configuration indoor localization through user collaboration. In: Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments, pp. 55–60.

Borrelli, A., Monti, C., Vari, M., Mazzenga, F., 2004. Channel models for IEEE 802.11 b indoor system design. In: 2004 IEEE International Conference on Communications, vol. 6, pp. 3701–3705.

Caso, G., De Nardis, L., 2015. On the applicability of multi-wall multi-floor propagation models to WiFi fingerprinting indoor positioning. In: Future Access Enablers of Ubiquitous and Intelligent Infrastructures, pp. 166–172.

Caso, G., De Nardis, L., 2017. Virtual and oriented WiFi fingerprinting indoor positioning based on multi-wall multi-floor propagation models. Mob. Netw. Appl. 22 (5), 825–833.

Caso, G., De Nardis, L., Di Benedetto, M.G., 2015a. A mixed approach to similarity metric selection in affinity propagation-based WiFi fingerprinting indoor positioning. Sensors 15 (11), 27692–27720.

Caso, G., De Nardis, L., Di Benedetto, M.G., 2015b. Frequentist inference for WiFi fingerprinting 3D indoor positioning. In: 2015 IEEE International Conference on Communication Workshop (ICCW), pp. 809–814.

Caso, G., De Nardis, L., Lemic, F., Handziski, V., Wolisz, A., Di Benedetto, M.G., 2016. ViFi: virtual fingerprinting WiFi-based indoor positioning via multi-wall multi-floor propagation model. ArXiv preprint arXiv:1611.09335.

Chintalapudi, K., Padmanabha Iyer, A., Padmanabhan, V.N., 2010. Indoor localization without the pain. In: Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking, pp. 173–184.

Damosso, E., 1999. COST ACTION 231: digital mobile radio towards future generation systems. Final Report, Tech. Rep., European Communities, EUR 18957.

Feng, C., Au, W.S.A., Valaee, S., Tan, Z., 2012. Received-signal-strength-based indoor positioning using compressive sensing. IEEE Trans. Mobile Comput. 11 (12), 1983–1993.

Frey, B.J., Dueck, D., 2007. Clustering by passing messages between data points. Science 315 (5814), 972–976.

Hernández, N., Ocaña, M., Alonso, J.M., Kim, E., 2017. Continuous space estimation: increasing WiFi-based indoor localization resolution without increasing the site-survey effort. Sensors 17 (1), 147.

Honkavirta, V., Perala, T., Ali-Loytty, S., Piché, R., 2009. A comparative survey of WLAN location fingerprinting methods. In: 6th Workshop on Positioning, Navigation and Communication, 2009. WPNC 2009, pp. 243–251.

Hossain, A.K.M.M., Van, H.N., Jin, Y., Soh, W.S., 2007. Indoor localization using multiple wireless technologies. In: IEEE International Conference on Mobile Adhoc and Sensor Systems, 2007. MASS 2007, pp. 1–8.

Kessel, M., Werner, M., 2011. SMARTPOS: accurate and precise indoor positioning on mobile phones. In: Proceedings of the First International Conference on Mobile Services, Resources, and Users, MOBILITY, pp. 158–163.

Kumar, S., Hegde, R.M., Trigoni, N., 2016. Gaussian process regression for fingerprinting based localization. Ad Hoc Netw. 51, 1–10.

Laoudias, C., Zeinalipour-Yazti, D., Panayiotou, C.G., 2013. Crowdsourced indoor localization for diverse devices through radiomap fusion. In: 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–7.

Le Dortz, N., Gain, F., Zetterberg, P., 2012. WiFi fingerprint indoor positioning system using probability distribution comparison. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2301–2304.

Liao, I.-E., Kao, K.-F., 2008. Enhancing the accuracy of WLAN-based location determination systems using predicted orientation information. Inform. Sci. 178 (4), 1049–1068.

Liu, H., Darabi, H., Banerjee, P., Liu, J., 2007. Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. C (Appl. Rev.) 37 (6), 1067–1080.

Marcus, P, Kessel, M., Werner, M., 2013. Dynamic nearest neighbors and online error estimation for SMARTPOS. Int. J. Adv. Internet Technol. 6 (1–2), 1–11.

Roos, T., Myllymäki, P., Tirri, H., Misikangas, P., Sievänen, J., 2002. A probabilistic approach to WLAN user location estimation. Int. J. Wirel. Inform. Netw. 9 (3), 155–164.

Shin, B., Lee, J.H., Lee, T., Kim, H.S., 2012. Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems. In: 2012 8th International Conference on Computing Technology and Information Management (ICCM), vol. 2, pp. 574–577.

Tian, Z., Tang, X., Zhou, M., Tan, Z., 2013. Fingerprint indoor positioning algorithm based on affinity propagation clustering. EURASIP J. Wirel. Commun. Netw. 2013 (1), 272.

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, Ó., Huerta, J., 2015. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Syst. Appl. 42 (23), 9263–9278.

Youssef, M., Agrawala, A., 2004. Handling samples correlation in the Horus system. In: INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 1023–1031.

Youssef, M.A., Agrawala, A., Shankar, A.U., 2003. WLAN location determination via clustering and probability distributions. In: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003 (PerCom 2003), pp. 143–150.

Yu, F., Jiang, M., Liang, J., Qin, X., Hu, M., Peng, T., Hu, X., 2014. 5 G WiFi signal-based indoor localization system using cluster k-nearest neighbor algorithm. Int. J. Distrib. Sens. Netw. 10 (12), 247525.