ViFi: Virtual Fingerprinting WiFi-based Indoor Positioning via Multi-Wall Multi-Floor Propagation Model

Giuseppe Caso, *Member, IEEE*, Luca De Nardis, *Member, IEEE*, Filip Lemic, *Member, IEEE*,
Vlado Handziski, *Member, IEEE*, Adam Wolisz, *Senior Member, IEEE*,
and Maria-Gabriella Di Benedetto, *Fellow, IEEE*

Abstract—Widespread adoption of indoor positioning systems based on WiFi fingerprinting is at present hindered by the large efforts required for measurements collection during the offline phase. Two approaches were recently proposed to address such issue: crowdsourcing and RSS radiomap prediction, based on either interpolation or propagation channel model fitting from a small set of measurements. RSS prediction promises better positioning accuracy when compared to crowdsourcing, but no systematic analysis of the impact of system parameters on positioning accuracy is available.
This paper fills this gap by introducing ViFi, an indoor positioning system that relies on RSS prediction based on Multi-Wall Multi-Floor (MWMF) propagation model to generate a discrete RSS radiomap (*virtual* fingerprints). Extensive experimental results, obtained in two different testbeds, show that ViFi outperforms virtual fingerprinting systems adopting simpler propagation models in terms of accuracy, and allows a sevenfold reduction in the number of measurements to be collected, while achieving the same accuracy of a traditional fingerprinting system deployed in the same environment. Finally, a set of guidelines for the implementation of ViFi in a generic environment, that saves the effort of collecting additional measurements for system testing and fine tuning, is proposed.

Index Terms—Indoor Positioning, WiFi Fingerprinting, Indoor Propagation Modeling, Multi-Wall Multi-Floor Model, Crowdsourcing.

◆

1 Introduction

INDOOR positioning and navigation is a market with expected size of USD 4 billions in 2019 [1], and both research community and industry are currently investing huge efforts in the search for a simple and yet reliable solution to determine the position of a mobile device in an indoor environment. Received Signal Strength (RSS) fingerprinting based on WiFi technology emerged as one of the most popular approaches for the implementation of Indoor Positioning Systems (IPSs) [2], [3], for two main reasons: 1) it allows to leverage existing communications infrastructure; 2) it can support almost any user device by simply installing an application, making it much easier to deploy than recent proposals, that may achieve submeter accuracy but still require specific hardware and/or software modifications to devices [4], [5].

WiFi fingerprinting traditionally operates in two phases. During the so-called offline phase, RSS values from WiFi Access Points (APs) detected in the environment (fingerprints) are collected at selected positions, referred to as Reference Points (RPs), in order to create a discrete RSS *radiomap* of the area of interest. Within the subsequent online phase, the location of target devices is estimated as a function of the positions of the RPs, that best match the RSS values measured by the devices. Accuracy and complexity of fingerprinting algorithms mainly depend on two issues: 1) proper definition of similarity metrics and estimation algorithms for selecting the best matching RPs to be used during the online phase, and 2) careful planning of the offline phase, particularly in terms of RP cardinality and positions, and of number of measurements collected at each RP [6], [7].

Regarding the online phase, deterministic and probabilistic k-Nearest Neighbors (kNN) and Weighted kNN (WkNN) algorithms are by far the most widely proposed and investigated: on one hand, deterministic algorithms are appealing and relatively easy to implement, because they take advantage of easily computable deterministic similarity metrics [8], [9]; on the other hand, probabilistic algorithms may improve the deterministic accuracy at the price of higher computational complexity and measurement efforts, due to the need of reliably estimating the RSS probability distributions in each RP from each WiFi AP [10], [11], [12].

The offline phase poses however major issues in large scale deployment of WiFi fingerprinting: the collection of measurements is increasingly time consuming with the area covered by the IPS, and requires on site measurement campaigns that potentially interfere with the activities that are usually carried out in the area. WiFi fingerprinting may therefore be difficult to deploy in cases such as:

security-sensitive facilities or prisons, characterized by restricted access areas, where measurements cannot be col-

- Giuseppe Caso, Luca De Nardis and Maria-Gabriella Di Benedetto are with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy.
  E-mail: caso@diet.uniroma1.it
  E-mail:{luca.denardis, mariagabriella.dibenedetto}@uniroma1.it
- Filip Lemic, Vlado Handziski and Adam Wolisz are with the Department of Telecommunication Systems, Telecommunication Networks Group (TKN), Technical University of Berlin, Berlin, Germany.
  E-mail: {lemic, handziski, wolisz}@tkn.tu-berlin.de

Manuscript received –; revised –.

lected;

emergency and diagnostic rooms in hospitals, where activities cannot be interrupted by measurement collection;
large shopping malls and skyscrapers spanning over tens or even hundred of floors, where the sheer size and extension make the collection efforts nonviable due to costs or time considerations.

Incidentally however, the above environments are some of the most appealing targets for an IPS system [13], [14], [15]. As a consequence, methods for simplifying the offline phase were proposed, and in particular:

RSS prediction - In this approach, most of the RSS values are predicted rather than measured. RSS prediction may be used either for the generation of virtual RPs[1], leading to *discrete* virtual fingerprinting, or for the evaluation of continuous RSS spatial distributions, that are used to infer the position, leading to *continuous* virtual fingerprinting. RSS prediction may be obtained by either *propagation modeling*, in which an empirical radio propagation model trained with a small set of measurements is used, or by *interpolation*, in which adjacent real RPs are interpolated.

Crowdsourcing - In this approach, devices that make use of the IPS contribute to the collection of location-dependent RSS samples [16], in either a voluntary or involuntary way [17]. Application of crowdsourcing entails a further challenge, that is heterogeneity, both in terms of location of RPs, as they are determined by the mobile device location, and of devices used for collection [18], [19], [20].

Both crowdsourcing and RSS prediction by interpolation require however on site access of the area covered by the IPS, and fail to meet the requirements posed by the scenarios previously identified.

Oppositely, in [21] it was shown that RSS prediction by propagation modeling can reliably predict RSS values in areas where no measurements are available, based on measurements taken in a different environment with similar propagation characteristics (e.g. different floors in the same building), forming the basis for ViFi, the WiFi IPS based on discrete virtual fingerprinting proposed and analyzed in this paper.

ViFi adopts RSS propagation modeling based for the offline phase, and combines it with a deterministic position estimation algorithm, with algorithm parameters automatically selected based on the settings used during the offline phase for the generation of virtual RPs. Design strategies for both offline and online phases were driven by the goal of limiting complexity: 1) the Multi-Wall Multi-Floor (MWMF) was selected among several empirical models, given its analytic simplicity [21], [22], [23], [24], [25], and 2) a WkNN algorithm was adopted during the fingerprinting online phase.

The proposed ViFi system was implemented and tested in two testbeds with different AP topologies and signal coverage characteristics, in order to obtain consolidated data supporting the proposed approach, and explore the possibility of introducing a set of guidelines for its seamless deployment in other testbeds.

----

1. *real* RPs indicate real fingerprints collected at the corresponding RP locations, while *virtual* RPs indicate virtual fingerprints generated as a function of the selected RSS prediction method and the corresponding RP locations.

The paper is organized as follows: Section 2 reviews related work in the field of RSS prediction, and identifies open issues and contributions of this work beyond the current state of the art. Section 3 contains the analytic foundations of the MWMF indoor propagation model, adopted for the generation of virtual RPs. The ViFi system is described in Section 4, where design choices related to both offline and online phases are discussed. The testbeds used for the experimental analysis are presented and compared in Section 5, with results being discussed in Section 6. Section 7 proposes the system implementation guidelines. Finally, Section 8 compares ViFi with traditional fingerprinting in terms of design requirements, positioning accuracy, and computational complexity, while Section 9 concludes the paper and highlights possible future research lines.

2　Virtual fingerprinting by RSS prediction

2.1　RSS Prediction by Interpolation

Both reduction and elimination of the offline phase have been pursued with RSS prediction by interpolation. Considering the reduction of the offline phase, discrete virtual fingerprinting with Inverse Distance Weighting (IDW) and Universal Kriging (UK) interpolation schemes was proposed in [8], while a linear regression approach was analyzed in [29]. Experimental results showed, in both cases, a positioning error decrease when compared with fingerprinting systems using a low amount of real RPs. Moreover, continuous virtual fingerprinting via Support Vector Regression (SVR) and Gaussian Processes (GP) was recently proposed in [30] and [31], respectively. In the first case, a slightly better accuracy with respect to a traditional kNN estimator was obtained; in then second case, a 30% accuracy improvement, with respect to the Horus system [11], was achieved.

Considering the elimination of the offline phase, [32] proposed a continuous virtual fingerprinting system using linear regression, relying on the capability by the APs to mutually measure the RSS of the beacon packets, characterized by an increase of 15% in the average positioning error compared to a traditional fingerprinting system.

2.2　RSS Prediction by Indoor Propagation Modeling

As in the case of interpolation, both reduction and elimination of the offline phase have been pursued with RSS prediction by propagation modeling. Considering the reduction of the offline phase, an empirical propagation model taking into account the effect of obstructing walls on the perceived RSS was proposed in the seminal work on WiFi fingerprinting [6]. The so-called Wall Attenuation Factor (WAF) was derived by averaging the differences between Line of Sight (LoS) and Non LoS (NLoS) measurements, with a known and variable amount of obstructing walls in the latter case. Once the WAF was evaluated, other propagation model parameters (path loss at a reference distance and path loss exponent, respectively) were computed via linear regression, and RSS values were predicted by using the resulting model. Experimental results showed an increase of the median positioning error in the order of 46% when such model was used in place of real measurements, calling

for the definition of a more accurate propagation model. In [22] the impact of using discrete virtual fingerprinting on the achievable positioning accuracy was comparatively analyzed: two empirical propagation models (log-distance and MWMF) and a semi-deterministic model (Motif) were used in the RSS prediction and virtual fingerprints generation. Positioning accuracy was then tested for deterministic NN and probabilistic Bayesian estimators, showing an increase of about 30% in terms of average positioning error for the log-distance model, and a decrease of about 10% for MWMF and Motif ones, with respect to the NN algorithm adopting real measurements. No analysis on the impact of the amount of both real and virtual RPs was however provided and, in addition, little detail was given on the set of propagation parameters used in the MWMF model. Recently, the possibility of generating reliable virtual RPs, through empirical fitting of a simple propagation model, was confirmed in [33], where a log-distance model was adopted, limiting the optimization to the path loss exponent. A more complex model was proposed in [34], foreseeing the sum of two exponentials, with a total set of four parameters to be estimated, but the model was not used to reduce the number of offline measurements.

Considering the elimination of the offline phase, a joint estimation of propagation parameters, distances and positions of both APs and user devices was proposed in [35], leading to an increase of positioning error by a factor of 2 or more when compared to a fingerprinting system. In general, however, offline phase-free systems have been proposed through the use of additional hardware or the modifications of off-the-shelf devices. In particular, the placement of WiFi sniffers at known locations was proposed in [36], with the goal of continuously measuring the RSS values from the APs in the area. A log-distance path loss model was then used in a continuous virtual fingerprinting approach. A similar system was proposed in [37], adopting however a more accurate ray tracing propagation model. In [38], an offline phase-free system named WILL was proposed, that combines temporal and spatial characteristics of indoor WiFi propagation with data provided by an accelerometer embedded in the device; experimental results showed an 86% room level accuracy in a typical office environment.

2.3 Open issues and proposed contribution

A first open issue emerging from the analysis of previous work is how to deal with the drawbacks related to complete removal the offline phase. Existing proposals either introduce modifications of off-the-shelf devices [32], [36], [37] and additional hardware such as inertial sensors [38], making their widespread adoption harder, or accept a significant increase in positioning error [35].

A second issue is related to the difficulty in achieving a satisfactory trade off between complexity and accuracy. Continuous virtual fingerprinting, in particular, is expected to slightly improve the positioning performance of a discrete virtual approach, although, in the literature, no clear evidence by direct comparison is provided. Furthermore, any performance gain would come at the price of an increase in system complexity [30], [31], [36], [37].

A third issue, specific to discrete virtual fingerprinting, is the

need for a precise knowledge on the impact of number and positions of virtual RPs on achievable positioning accuracy [8], [29], [22]. The relation between offline settings and online algorithm parameters is also unclear, leading to a case by case choice rather than selection based on significant system parameters. This is the case, in particular, for k in kNN/WkNN algorithms [3], [6].

A last issue regards the need for sets of measurements acquired in a controlled and reliable fashion, upon which all proposals reviewed in Sections 2.2 and 2.1 rely. The use of crowdsourced measurements for RSS prediction could further reduce the efforts required for the offline phase, but this possibility is all but unexplored in the literature, and its impact on positioning accuracy is unknown.

Regarding the first issue, the contribution provided in this paper is to quantitatively determine the minimum amount of measurements required to achieve an accuracy comparable to a real fingerprinting system, without the need for dedicated and/or modified infrastructure.

Regarding the second issue, the virtual fingerprinting system proposed in this work, ViFi, adopts the MWMF model, that is characterized by a complexity comparable to the simpler models proposed in [6], [33], [36], leading however to an accuracy similar to a real fingerprinting system deployed in the same environment.

As for the third issue, with respect to [8], [29], the present work proposes a more detailed analysis of the impact of densities and positions of both real and virtual fingerprints on: a) reliability in generating virtual fingerprints by MWMF, and b) positioning accuracy. In addition, with respect to [22], a more complete analysis of the MWMF model is provided. Furthermore, this paper proposes guidelines for the implementation and setup of a virtual fingerprinting system by indicating: a) the required spatial densities of both real and virtual RPs, b) the strategies to select and place such RPs, and c) the value of k, guaranteeing satisfactory performance without requiring a training phase.

Finally, with respect to the fourth issue, preliminary results are provided on the combination of virtual fingerprinting and crowdsourcing measurements, offering a framework on which further investigations may develop.

3 Multi-Wall Multi-Floor Propagation Models

Multi-Wall Multi-Floor models [23] emerged among empirical narrow-band models as an appealing solution for indoor propagation modeling, due to the good trade off they provide between analytic simplicity and path loss modeling accuracy [24]. MWMF models take into account objects, that obstruct signal propagation in an indoor wireless link, leading to the following path loss model [23]:

$$\mathsf{PL}_{\mathsf{MWMF}} = \mathsf{PL}_{\mathsf{OS}} + A_{\mathsf{MWMF}} \quad [\mathsf{dB}], \qquad (1)$$

where the $\mathsf{PL}_{\mathsf{OS}}$ term models the path loss over the Tx-Rx distance d, while A_{MWMF} models the additional loss due to obstructing obstacles. The $\mathsf{PL}_{\mathsf{OS}}$ term is defined according to the One Slope model, as follows [23]:

$$\mathsf{PL}_{\mathsf{OS}}(d, \gamma) = l_0 + 10\gamma \log(d) \quad [\mathsf{dB}], \qquad (2)$$

with l_0 modeling the $d = 1$ m reference path loss, while γ is the path loss exponent (for free space conditions, $\gamma = 2$ and $l_0 \approx 40.22$ dB @ 2.45 GHz [24]). When, in a more general case, Tx and Rx are located on different floors, A_{MWMF} is given by [23]:

$$A_{\text{MWMF}} = l_c + \sum_{n=1}^{N_{\text{obj}}} \sum_{i=1}^{I_n} N_{n,i} l_{n,i} + N_{\text{f}}^{\left[\frac{N_{\text{f}}+2}{N_{\text{f}}+1} - b \right]} l_{\text{f}} \quad [\text{dB}], \quad (3)$$

the parameters of which are described in TABLE 1.

TABLE 1: A_{MWMF} parameters description.

Parameter	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

The use of the MWMF model requires an initial set of M measurements, and also for each measurement ($m = 1, 2, \ldots, M$) the information regarding cardinality, type and positions of objects obstructing the m-th 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 $\{\mathcal{S}\}$ characterizing the model. In this paper, a least square fitting procedure that minimizes the difference between RSS measurements and predictions was adopted to estimate $\{\mathcal{S}\}$. The optimal $\{\mathcal{S}\}_{\text{opt}}$ is thus obtained as follows:

$$\{\mathcal{S}\}_{\text{opt}} = \underset{\{\mathcal{S}\}}{\arg\min} \left\{ \sum_{m=1}^{M} |\text{RSS}_m - \hat{\text{RSS}}_m|^2 \right\}, \quad (4)$$

where, for the m-th available measurement, RSS_m and $\hat{\text{RSS}}_m$ are the actual vs. the predicted RSS values at Rx, when considering a Tx emitting a known Effective Isotropic Radiated Power (EIRP) $\text{W}_{\text{TX}}^{\text{EIRP}}$ at distance d_m. $\hat{\text{RSS}}_m$ is computed as follows:

$$\hat{\text{RSS}}_m = \text{W}_{\text{TX}}^{\text{EIRP}} - \text{PL}_{\text{MWMF}}(d_m, \{\mathcal{T}_m\}, \{\mathcal{S}\}). \quad (5)$$

The propagation parameters included in $\{\mathcal{S}\}$ may differ from one MWMF model to the other, and can include parameters characterizing both PL_{OS} and A_{MWMF}; the set adopted in this work is defined in Section 5.3.

4 ViFi System Model

4.1 Offline Phase

ViFi uses the MWMF model for the generation of virtual RP fingerprints. Given a set of L WiFi APs, in known positions, measurements in a set of N^{r} real RPs are first collected, so that a $L \times 1$ RSS fingerprint $\boldsymbol{s}_{n_1}$ is associated with the n_1-th RP. The generic $\boldsymbol{s}_{n_1}$ component, denoted by s_{l,n_1}, contains the RSS received at the n_1-th RP from the l-th AP, obtained by averaging $q > 1$ measurements in order

to counteract channel variability. The selection of q is a compromise between system accuracy vs. time and effort devoted to measurements. Since only a subset of the L APs may be detected at the generic RP the $\boldsymbol{s}_{n_1}$ components of undetected APs are set to a predefined value reflecting lack of detection.

The MWMF model is then calibrated on the set of real fingerprints, and used for the generation of N^{v} virtual fingerprints associated with 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_2-th virtual RP from the l-th AP.

4.2 Online Phase

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

$$\hat{\boldsymbol{p}}_i = \frac{\sum_{n=1}^{k}(sim_{n,i})\boldsymbol{p}_n}{\sum_{n=1}^{k} sim_{n,i}}, \quad (6)$$

where $\boldsymbol{p}_n = (x_n, y_n, z_n)$ is the position of the n-th RP in a 3D coordinate system, and $\hat{\boldsymbol{p}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$ is the estimated position of the target device.

$sim_{n,i}$ can be any deterministic metric computable in the RSS space between vectors $\boldsymbol{s}_n$ and $\boldsymbol{s}_i$. A popular choice is the inverse Minkowski distance of order o, defined as follows:

$$sim_{n,i} = [\mathcal{D}_{n,i}^o]^{-1} = \left[\left(\sum_{l=1}^{L} |s_{l,i} - s_{l,n}|^o \right)^{\frac{1}{o}} \right]^{-1}, \quad o \geq 1. \quad (7)$$

Typical orders are $o = 1$ (inverse Manhattan distance) and $o = 2$ (inverse Euclidean distance, used in ViFi). Similarity metrics based on modified versions of the inner product between RSS vectors have also been proposed [39], [40].

4.3 Offline Phase Implementation Options

The ViFi offline phase must address two main issues: a) how to use the real RPs, and b) how to determine the number and positions of virtual RPs, as discussed below.

4.3.1 Handling real RPs

Both cardinality and positions of real RPs are expected to affect the generation of virtual RPs and the resulting ViFi positioning accuracy. Given N^{r} real RPs regularly spaced over a grid in $\mathcal{A}$ with area $|\mathcal{A}|$, the spatial RP density is:

$$d^{\text{r}} = \frac{N^{\text{r}}}{|\mathcal{A}|}. \quad (8)$$

In [21], [25], several strategies were proposed for selecting the propagation parameters for the MWMF model;

among them, in this paper the following two strategies, in which measurements collected at all RPs are jointly used to estimate propagation parameters, were adopted:

Strategy I: Environment Fitting - the set of measurements at all RPs collected from all APs are used in a global optimization procedure, leading to a common set of propagation parameters used for all APs in $\mathcal{A}$. The underlying assumption in Strategy I is that one set of parameters can globally characterize the environment.

Strategy II: Specific AP Fitting - the set of measurements at all RPs referring to a specific AP are used to estimate propagation parameters for that AP. The procedure leads thus to a different set of propagation parameters for each AP.

4.3.2 Handling virtual RPs

In analogy to Equation (8), denoting with N^{v} the number of virtual RPs to be generated, the virtual RPs spatial density be defined as follows:

$$d^{\text{v}} = \frac{N^{\text{v}}}{|\mathcal{A}|}. \tag{9}$$

Moreover, positions of virtual RPs in the area can be freely defined: possible options include placement on a grid, as a natural extension of the placement scheme used for real RPs, and random placement.

4.4 Online Phase Implementation Options

The WkNN algorithm foresees the selection of two parameters: parameter k and weighting function $sim_{n,i}$. In ViFi the weighting function $sim_{n,i}$ is set to the inverse Euclidean distance while parameter k is selected according to a new procedure. Previous work addressed the impact of k on positioning accuracy of deterministic WkNN algorithms in real fingerprinting systems, and two approaches emerged: a) a dynamic k selection i.e. that, however, increases complexity without guaranteeing the optimal performance in all cases [40], [41], and b) a static k selection, that minimizes the average positioning error over a set of Test Points. The parameter k that minimizes the average error is referred to in the following as k_{opt}. As shown in [3], [6], [7], [8], [39], [41] in real fingerprinting systems k_{opt} typically lies in the range between 2 and 10.

This paper investigates the determination of k_{opt} on the basis of system parameters, moving from the assumption that k_{opt} depends in particular on the density of real and virtual RPs, and can be thus determined as $k_{\text{opt}} = \mathsf{f}(d^{\text{r}}, d^{\text{v}})$. In this work a linear law is proposed for $\mathsf{f}(.,.)$, leading to an approximate k_{opt} value, k_{est}, as follows:

$$k_{\text{opt}} \approx k_{\text{est}} = \lceil \alpha(d^{\text{r}} + d^{\text{v}})|\mathcal{A}| \rceil, \quad \alpha \ll 1. \tag{10}$$

Equation (10) assumes the same linear dependency of k_{opt} on both d^{r} and d^{v}, although virtual fingerprints are different from real ones, since they show a high spatial correlation, with variation of RSS values between two fingerprints directly related to the physical distance between the corresponding RPs, without the abrupt changes introduced in real fingerprints by measurement errors and channel fading. Two reasons justify this choice: first, there is no

model in the literature providing the value of k_{opt} as a function of the density of real fingerprints; second, virtual fingerprinting only makes sense if the density of real RPs is much smaller compared to the density of virtual RPs, thus leading to a low impact of d^{r} on k_{opt}.

In general, the values of α, d^{r} and d^{v} to be used in Equation (10) will be environment-specific, and will still require a testing phase to determine their exact values. If however a set of values for these parameters valid across different testbeds environments can be determined, this will allow to put the system in operation without the need for a testing phase. This possibility is investigated later in this work, first by assessing the validity of Equation (10) with the procedure described in Section 5.5, and then by determining and comparing the values of α, d^{r} and d^{v} that lead to the best approximation of k_{opt} in SPinV and TWIST testbeds, adopterd in this work for the experimental analysis.

5 Experimental Analysis Setup

This section describes the SPinV and TWIST testbeds, implemented at Sapienza University of Rome and Technical University of Berlin, respectively. Next, the procedures defined for the experimental analysis of the ViFi system are introduced.

5.1 SPinV

Supporting People indoor: a navigation Venture (SPinV) is the indoor positioning testbed implemented at the DIET Department of Sapienza University of Rome. SPinV is deployed in an office environment and covers two floors with an area of approximately 42×12 m^2 each. $L_1 = 6$ WiFi APs working @ 2.4 GHz and 5 GHz, and $L_2 = 7$ WiFi APs working @ 2.4 GHz, with a beacon transmission period of $T_{\text{b}} = 100$ ms and a transmit power $W_{\text{TX}}^{\text{EIRP}} = 20$ dBm, are placed at known positions at the 1-st and 2-nd floor, respectively. In this work, the SPinV 2-nd floor has been adopted as evaluation environment $\mathcal{A}$, and the APs on this floor have been considered in the fingerprinting measurement campaign, so that $L = L_2$. During the offline phase, $N^{\text{r,tot}} = 72$ RPs were selected for fingerprints collection on a regular grid within $\mathcal{A}$; fingerprints were also collected in a set of $N^{\text{t}} = 31$ Test Points (TPs) randomly distributed over $\mathcal{A}$.

Two different measurement campaigns were carried out in the SPinV testbed, each corresponding to a different application scenario:

Controlled scenario - All measurements in both RPs and TPs were carried out during weekend afternoons in order to mitigate the effect of radio interference and variation of propagation conditions due to human presence. A MacBook Pro equipped with an AirPort Extreme Network Interface Card and placed on a wooden platform was used in the campaign in order to rule out the impact of the human body presence on measured RSS. Both RPs and TPs fingerprints were obtained as the average of $q = 50$ scans at each location, in order to remove as much as possible fluctuations due to channel and measurement variability.

Crowdsourcing-like scenario - In this scenario, all measurements were performed during weekdays, by human

surveyors equipped with Android-based Samsung tablets. RPs fingerprints were obtained by averaging $q = 5$ measurements, while TPs fingerprints include a single measurement for each location.

A specific characteristic of the SPinV testbed is the location of APs. On both floors, APs are in fact placed in the false ceiling of the central corridor, as shown in Figure 1. Furthermore, not all APs cover the entire floor. SPinV can be thus considered an example of sub-optimal environment in terms of signal coverage and APs topology, leading to low spatial diversity.

5.2 TWIST

The TKN Wireless Indoor Sensor network Testbed (TWIST) is the indoor positioning testbed implemented at the TKN headquarter at TUB, in the context of the EVARILOS Project [26], [27]. TWIST is deployed in an office environment and covers one floor with an area of about 30×15 m^2. Within this environment, $L = 4$ dedicated WiFi APs were placed at known positions and configured to operate @ 2.4 GHz with a beacon transmission period of $T_\text{b} = 100$ ms and a transmit power $\text{W}_\text{TX}^\text{EIRP} = 20$ dBm. The entire floor was adopted as evaluation environment $\mathcal{A}$, and all the L APs were included in the fingerprinting measurement campaign. During the fingerprinting offline phase, $N^\text{r,tot} = 41$ RPs were identified on a regular grid and measured within $\mathcal{A}$; an additional set of $N^\text{t} = 80$ uniformly distributed TPs was collected.

In the case of the TWIST testbed, all measurements were taken with the same settings defined for the Controlled scenario in SPinV, and no Crowdsourcing-like scenario was considered.

Oppositely to SPinV, APs in TWIST are approximately placed at the four corners of the coverage area, as shown in Figure 2, to cover the entire area. TWIST can be thus considered as an optimal environment in terms of signal coverage and AP topology, characterized by high spatial diversity.

5.3 Procedure for the Analysis of the MWMF Accuracy in Virtual RPs Generation

In agreement with the approach proposed in [21], [25], the following procedure was adopted for the analysis of the reliability of the MWMF model in generating virtual RPs:

1) A parameter ρ was defined, in order to determine the number of RPs, N^r, used for the model fitting reported in Equation (4), out of the total number of RPs, $N^\text{r,tot}$, so that $N^\text{r} = \lceil \rho N^\text{r,tot} \rceil$. The modeling accuracy in estimating the propagation parameters and generating virtual RPs was then evaluated as a function of ρ. Figure 1 shows the positions of the $N^\text{r,tot}$ RPs, of the APs and of the N^r RPs selected for model fitting as a function of ρ in SPinV, while Figure 2 provides the same information for TWIST.

2) The model fitting procedure defined by Equation (4), introduced in Section 3, was carried out to evaluate propagation parameters to be used in the MWMF model for each of the two selection strategies introduced in Section 4.3.1. TABLE 2 defines the set $\{\mathcal{S}\}$ of propagation parameters included in the MWMF model
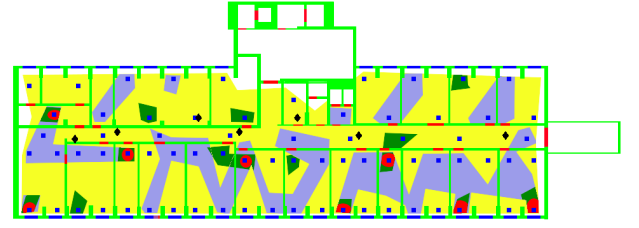

Fig. 1: Position of real RPs (blue squares) and APs (black diamonds) in the SPinV testbed. Areas of different colors highlight selected RPs when $\rho = 0.1$ (red), $\rho = 0.2$ (red + green), $\rho = 0.5$ (red + green + purple), $\rho = 1$ (all colors).
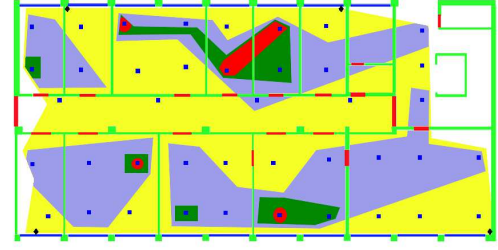

Fig. 2: Position of real RPs (blue squares) and APs (black diamonds) in the TWIST testbed. Areas of different colors highlight selected RPs when $\rho = 0.1$ (red), $\rho = 0.2$ (red + green), $\rho = 0.5$ (red + green + purple), $\rho = 1$ (all colors).

fitting procedure, and the settings for the remaining parameters in Equation (4).

TABLE 2: Model parameters setting.

Parameter	Setting
N_obj	2 (walls, doors)
I_walls	1
I_doors	1
N_f	0
$\{\mathcal{S}\}$	$\{\gamma, l_c, l_\text{wall}, l_\text{door}\}$

3) For each selection strategy, RSS values were predicted, using the MWMF model, in the same locations where the $N^\text{r,tot}$ RPs were originally collected. Next, the accuracy of the model was evaluated by defining the prediction error as follows:

$$\delta_{l,n}(\rho) = |s_{l,n} - \hat{s}_{l,n}(\rho)|, \tag{11}$$

where $\hat{s}_{l,n}(\rho)$ is the predicted RSS for the generic (AP$_l$, RP$_n$) pair, obtained by using a set of $N^\text{r} = \lceil \rho N^\text{r,tot} \rceil$ RPs in the model fitting procedure, while $s_{l,n}$ is the measured RSS for the same pair, measured as explained in Section 4.1. Assuming the generic $\delta_{l,n}(\rho)$ value as a sample of a random variable $\delta_l(\rho)$ related to the l-th AP, the cumulative distribution function (CDF) of $\delta_l(\rho)$ and the average error $\bar{\delta}_l(\rho) = \frac{\sum_{n=1}^{N^\text{r,tot}} \delta_{l,n}(\rho)}{N^\text{r,tot}}$ were also evaluated.

Results of the analysis are presented in Section 6.1 and 6.3 for Controlled and Crowdsourcing-like scenarios, respectively.

5.4 Procedure for the Analysis of the Impact of Virtual RPs on Positioning Accuracy

The performance of ViFi was evaluated as a function of: a) the density d^r and the selection strategy of real RPs, b) the density d^v and placement of virtual RPs, and c) the k parameter in the WkNN algorithm. The analysis was carried out by computing the positioning error $\epsilon_i(d^r, d^v, k)$ for each TP i $(i = 1, 2, \ldots, N^t)$ as follows:

$$\epsilon_i(d^r, d^v, k) = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2}, \quad (12)$$

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 positions of the i-th target device, respectively. Note that dependence of $\hat{\boldsymbol{p}}_i$ components on the $\{d^r, d^v, k\}$ set was omitted in Equation (12) for the sake of readability. As in the case of the prediction error $\delta_l(\rho)$, the CDF of positioning error $\epsilon(d^r, d^v, k)$ and the average positioning error $\bar{\epsilon}(d^r, d^v, k) = \frac{\sum_{i=1}^{N^t} \epsilon_i(d^r, d^v, k)}{N^t}$ have been evaluated.

Furthermore, a performance indicator, referred to as *virtualization* gain $\mathcal{G}(d^r, d^v, k)$, was introduced to measure the gain in accuracy achieved when using a density of virtual RPs $d^v > 0$. The virtualization gain $\mathcal{G}(d^r, d^v, k)$ is defined as follows:

$$\mathcal{G}(d^r, d^v, k) = \frac{\bar{\epsilon}(d^r, 0, k)}{\bar{\epsilon}(d^r, d^v, k)}, \quad d^v > 0. \quad (13)$$

Results of the analysis are presented in Section 6.2 and 6.3 for Controlled and Crowdsourcing-like scenarios, respectively.

5.5 Procedure for testing the validity of the k_{est} model

The following procedure was applied in order to assess the validity of Equation (10) in both SPinV and TWIST testbeds:

- Given a generic (d^r, d^v) combination, k_{opt} and in turn $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ were evaluated;
- For each $\alpha \in [\alpha_{\text{min}} : \alpha_{\text{max}}]$, $k_{\text{est}}(\alpha)$ was evaluated following Equation (10), and the corresponding average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{est}}(\alpha))$ was determined;
- The difference between $\bar{\epsilon}(d^r, d^v, k_{\text{est}}(\alpha))$ and $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ was evaluated and denoted with $\beta(d^r, d^v, \alpha)$:

$$\beta(d^r, d^v, \alpha) = \bar{\epsilon}(d^r, d^v, k_{\text{est}}(\alpha)) - \bar{\epsilon}(d^r, d^v, k_{\text{opt}}) \quad (14)$$

Numerical results of the above procedure for both SPinV and TWIST testbeds are reported in Section 6.2.4.

TABLE 3 presents the values for the parameters introduced in the procedures defined in the current section, as well as in Sections 5.3 and 5.4.

6 Experimental Results and Discussion

6.1 Controlled Scenario: Accuracy in Generation of Virtual RPs

The MWMF model adopted in ViFi for the generation of virtual RPs was analyzed and compared against the OS

TABLE 3: Experimental parameters setting.

Parameter	SPinV	TWIST
$\{\rho\}$	\{0.1, 0.2, 0.5, 1\}	
$\{d^r\}$	\{0.02, 0.03, 0.07, 0.14\}	\{0.01, 0.02, 0.05, 0.09\}
$\{N^r\}$	\{8, 15, 36, 72\}	\{5, 9, 21, 41\}
$\{d^v\}$	\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}	
$\{N^v\}$	\{6, 26, 51, 252, 504, 2520, 5040\}	\{5, 23, 45, 225, 450, 2250, 4500\}
$\{\alpha_{\text{min}}, \alpha_{\text{max}}\}$	\{0.01, 0.25\}	

model proposed in [22] and [33].

Figures 3 and 4 show the CDFs of the prediction error $\delta_l(\rho)$ for selection Strategy I (Figures 3a and 4a) and Strategy II (Figures 3b and 4b) for a reference AP within the SPinV and TWIST testbeds. Results show that, for both strategies and testbeds, slightly different $\delta_l(\rho)$ errors are obtained as ρ increases from 0.1 to 1. Strategies perform similarly for the reference AP in the SPinV environment, with a slight performance improvement given by Strategy II, while, within the TWIST environment, Strategy I allows a faster convergence to low values of the prediction error $\delta_l(\rho)$ with respect to Strategy II.
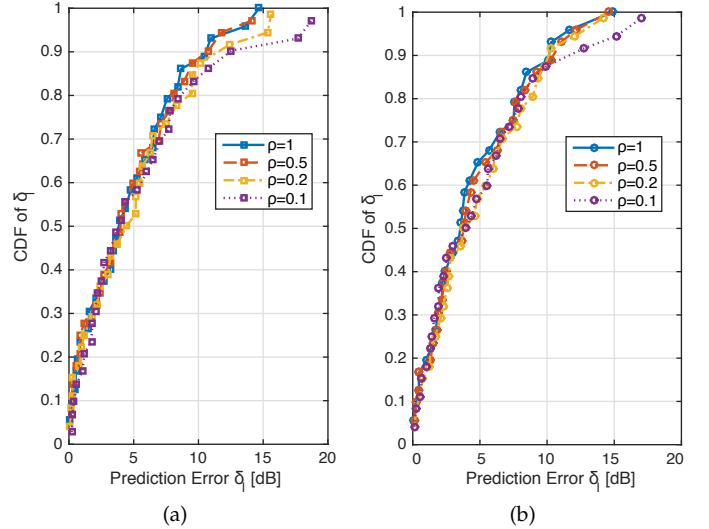

Fig. 3: Cumulative Distribution Function of prediction error $\delta_l(\rho)$ for a reference AP in SPinV ((a): Strategy I, (b): Strategy II.)

Results suggest that, when a relatively large value of ρ (and in turn of N^r and d^r) is used, Strategy II outperforms Strategy I. Oppositely, when a relatively low ρ value is used, Strategy I achieves slightly better performance than Strategy II, since it combines measurements from different APs, taking advantage of environment homogeneity. This is confirmed by Figure 5, showing the average prediction error $\bar{\delta}(\rho)$, obtained by averaging all the $\bar{\delta}_l(\rho)$ values over the L APs, as a function of ρ (in addition to the ρ values in TABLE 3, values of $\bar{\delta}(\rho)$ obtained for $\rho = 0.05$ are also shown), in SPinV and TWIST, respectively. The Figure also presents results obtained using the OS model proposed in [22], [33]. Results show that Strategy II is the optimal selection strategy in SPinV, since the initial amount of real RPs is large enough to guarantee a better RSS prediction when specific propagation parameters are associated to different APs. Oppositely,

Strategy I provides a better RSS prediction in the TWIST testbed for low values of ρ, taking advantage in particular of the symmetry in the AP placement in TWIST. Figure 5 also shows that the proposed MWMF model outperforms the OS model, with a decrease in the average prediction error of up to 4 dBs in SPinV and 2 dBs in TWIST. Results presented so
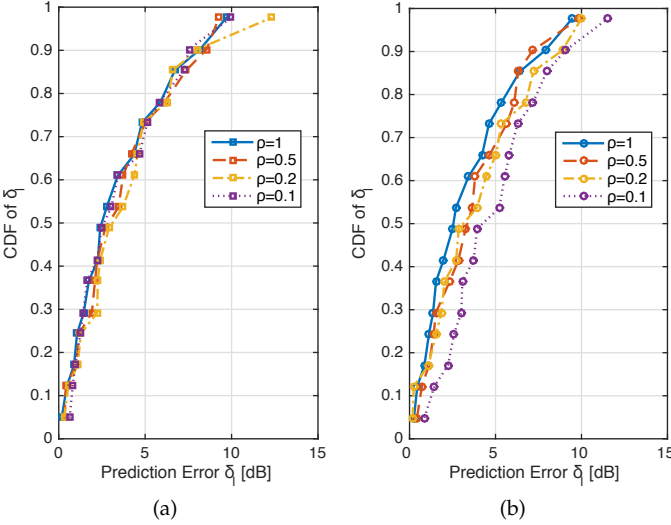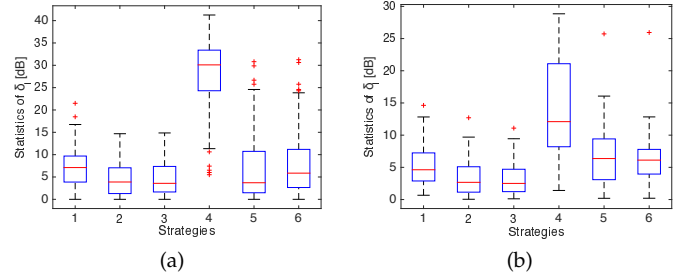

(a)

(b)

Fig. 6: Statistics of $\delta_l(\rho)$ for a reference AP and $\rho = 1$ in SPinV (a) vs. TWIST (b) (1: MWMF-No Fit; 2: MWMF-Strategy I; 3: MWMF-Strategy II; 4: OS-No Fit; 5: OS-Strategy I; 6: OS-Strategy II).

and possible outliers. Results show that the prediction error significantly increases for both MWMF and OS models, and in both testbeds, when no site-specific model fitting is carried out: measurements can be thus significantly reduced but not totally avoided, because of the empirical nature of the model.

6.2 Controlled Scenario: Positioning Accuracy

In this section, experimental results on the positioning accuracy of ViFi are presented for both testbeds in the Controlled scenario. Next, results of the application of the empirical procedure presented in Section 5.5 for estimating the value of k_{opt} in the ViFi online phase are reported.

6.2.1 Impact of d^r

Figures 7 and 8 show the impact of the density of real RPs on the average positioning accuracy, when no virtual RPs are introduced, for SPinV vs. TWIST testbeds.

Figures 7a and 8a report the average positioning error $\bar{\epsilon}(d^r, 0, k)$ as a function of k in SPinV vs. TWIST. Results show that the positioning error decreses as d^r increases. The decrease of $\bar{\epsilon}(d^r, 0, k)$ is less and less significant as d^r increases, suggesting the presence of a lower bound, in agreement with [6] and others, possibly due to the inherent measurement error in the collection of real RPs. Results also show that a different lower bound for the error is reached in the two testbeds, about 3 m for SPinV vs. 2 m for TWIST, and indicate that optimal AP placement positively affects the system performance, in agreement with [42].

Figures 7b and 8b present the boxplot of the positioning error $\epsilon(d^r, 0, k_{opt})$ as a function of d^r. Results confirm that in real fingerprinting systems the value of k_{opt} spans in the $[2, 10]$ range, as observed in previous works [3], [6], [7], [8], [39], [41].

6.2.2 Impact of real RPs selection strategies and virtual RPs placement

Figure 9 reports the average positioning error $\bar{\epsilon}(d^r, d^v_{max}, k_{opt})$ for Strategy I and II, and d^r as in TABLE 3. Results are in agreement with Figure 5: within the SPinV environment, Strategy II outperforms Strategy I, and strategies performance approach when $d^r = d^r_{min}$; on the contrary, within the TWIST environment, Strategy I leads to better positioning accuracy, and strategies performance

(a)

(b)

Fig. 4: Cumulative Distribution Function of prediction error $\delta_l(\rho)$ for a reference AP in TWIST ((a): Strategy I, (b): Strategy II.)

far highlight that few, uniformly distributed, measurements are in most cases sufficient to obtain a reliable estimation of the propagation parameters and generation of virtual RPs, supporting the approach proposed in ViFi.

One might ask: are measurements required at all? The question is answered by results shown in Figure 6, presenting the statistics of $\bar{\delta}_l(\rho)$ for a reference AP and $\rho = 1$ for the MWMF model vs. the OS model in [22], [33]. Each model was fitted according to Strategy I, Strategy II and to a baseline *No Fit* strategy in which no fitting is carried out. For the *No Fit* strategy, the RSS predictions were obtained in MWMF by using propagation parameters estimated for a totally different area $\mathcal{A}'$, reported in [24]; in the the OS model the settings $\gamma = 2$, $l_0 = 20$ were adopted. Note that in Figure 6 and in following ones error statistics are summarized in the form of boxplot diagrams showing minimum, maximum and median values, 25-th and 75-th percentiles,

(a)

(b)

Fig. 5: Average prediction error $\bar{\delta}(\rho)$ as a function of ρ for Strategies I and II in SPinV (a) vs. TWIST (b) (Continuous lines: proposed MWMF model; dashed-dotted lines: OS model proposed in [22], [33].)
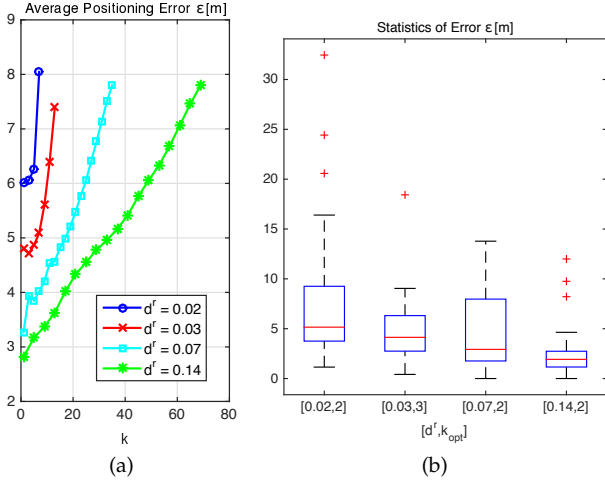
Fig. 7: Average positioning error $\bar{\epsilon}(d^r, 0, k)$ as a function of k for d^r as in TABLE 3 (a) and statistics of $\epsilon(d^r, 0, k_{opt})$ (b) in SPinV.
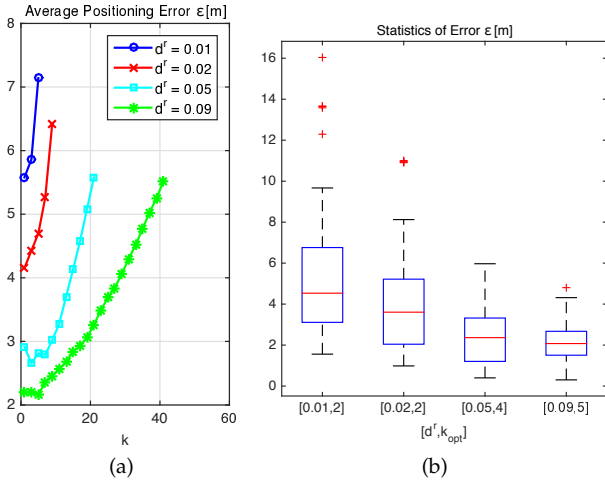

Fig. 8: Average positioning error $\bar{\epsilon}(d^r, 0, k)$ as a function of k for d^r as in TABLE 3 (a), and statistics of $\epsilon(d^r, 0, k_{opt})$ (b) in TWIST.

approach when $d^r = d^r_{max}$. In conclusion one can observe that: 1) the selection of the optimal strategy depends on the number of real RPs, and 2) a direct relationship exists between the average prediction error $\bar{\delta}(\rho)$ and the average positioning error $\bar{\epsilon}(d^r, d^v, k)$: the strategy minimizing the average prediction error $\bar{\delta}(\rho)$ also minimizes the average positioning error, and should thus be selected.

As for the placement of virtual RPs, both grid and random placements were investigated, according to the discussion in Section 4.3.2, with results showing no significant difference between the two placements. As a consequence, a grid placement was adopted in all experimental results shown throughout the paper.

6.2.3 Impact of d^v

In analogy with Figures 7 and 8 presented in Section 6.2.1, Figures 10 and 11 show the average positioning error $\bar{\epsilon}(d^r_{min}, d^v, k)$ as a function of k (Figures 10a and 11a) and the boxplot of the positioning error $\epsilon(d^r_{min}, d^v, k_{opt})$ (Figures 10b
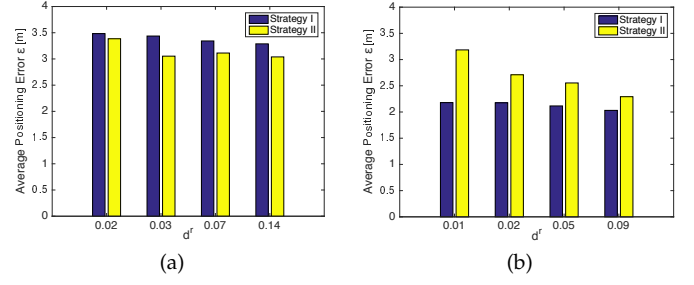

Fig. 9: Average positioning error $\bar{\epsilon}(d^r, d^v_{max}, k_{opt})$, for d^r as in TABLE 3, and for selection Strategy I and II, in SPinV (a) vs. TWIST (b) (Blue bars: Strategy I; yellow bars: Strategy II).

and 11b) in SPinV and TWIST testbeds, respectively. Results show that, when d^r_{min} is used, the introduction of virtual RPs significantly reduces the positioning error with respect to case of $d^v = 0$. Results are generalized and extended in Figure 12, showing the virtualization gain $\mathcal{G}(d^r, d^v, k_{opt})$, as defined in Equation (13), for both testbeds. Results show that the introduction of virtual RPs significantly improves positioning accuracy for low d^r values as d^v increases, while the advantage of introducing virtual RPs becomes negligible when a relatively large number of real RPs is available (high d^r values).
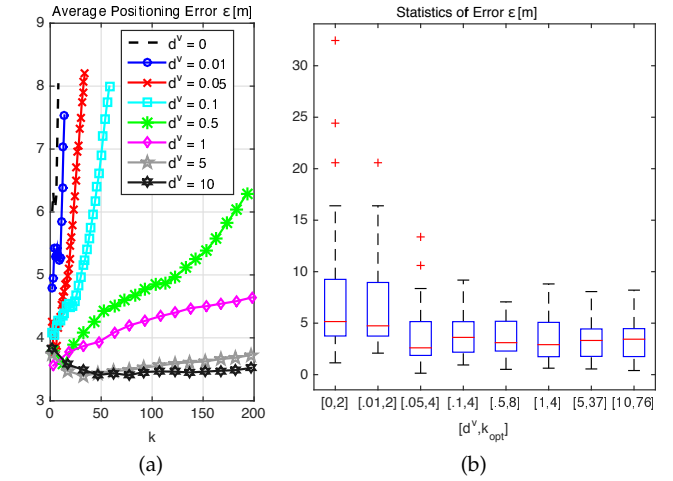

Fig. 10: Average positioning error $\bar{\epsilon}(d^r_{min}, d^v, k)$ as a function of k, for d^v as in TABLE 3 (a), and statistics of $\epsilon(d^r_{min}, d^v, k_{opt})$ (b) in SPinV.

Low d^r values were thus selected for further investigation, with results presented in Figure 13 and Figure 14, showing the average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ as a function of d^v for the two lowest values of d^r in SPinV vs. TWIST. Results are presented for both the MWMF model used in ViFi and the OS model of [22] and [33], as well as for a real fingerprinting system, with $d^r = d^r_{min}$ and $d^r = d^r_{max}$ as upper and lower bounds for positioning error, respectively. Results indicate that in both testbeds the minimum average positioning error is obtained for $d^v = d^v_{max}$, with the MWMF model adopted in ViFi leading to consistently better accuracy with respect to the OS model. Interestingly, results also show that the error measured with the MWMF model converges in all cases to $\bar{\epsilon}(d^r_{max}, 0, k_{opt})$ (with the only
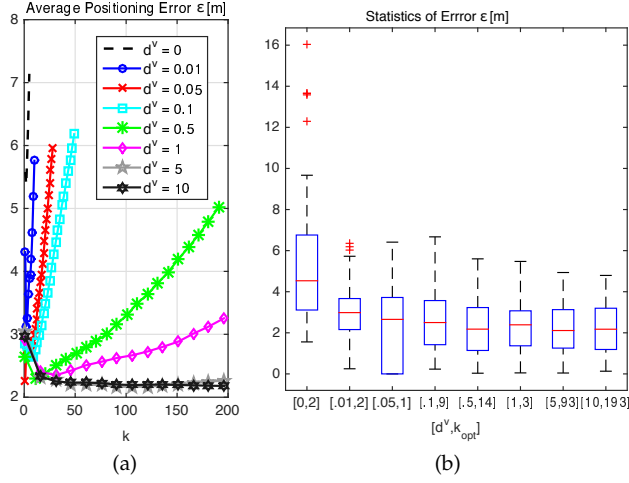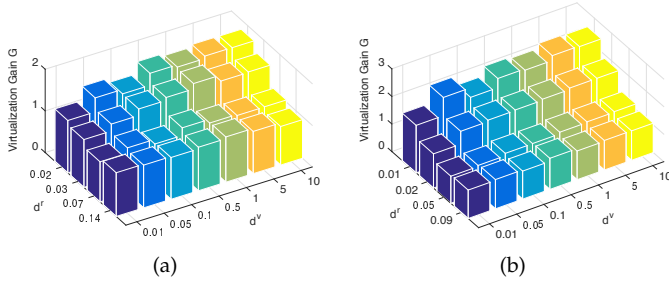
(a)

(b)

Fig. 11: Average positioning error $\bar{\epsilon}(d^r_{min}, d^v, k)$ as a function of k, for d^v as in TABLE 3 (a), and statistics of $\epsilon(d^r_{min}, d^v, k_{opt})$ (b) in TWIST.

(a)

(b)

Fig. 12: Virtualization gain $\mathcal{G}(d^r, d^v, k)$ as a function of d^r and d^v, for $k = k_{opt}$, in SPinV (a) vs. TWIST (b).

exception of $\bar{\epsilon}(d^r_{min}, d^v_{max}, k_{opt})$ in the SPinV testbed, about 60 cm above $\bar{\epsilon}(d^r_{max}, 0, k_{opt})$).

The results lead therefore to the conclusions that a) the introduction of virtual fingerprints in ViFi effectively compensates the reduction of real ones with negligible effects on positioning accuracy, allowing to significantly reduce time required for the offline phase, and b) the adoption of the MWMF model is instrumental in achieving this result.

6.2.4 Empirical Derivation of k_{est}

The value of α guaranteeing a reliable estimation of k_{opt} was determined, for both environments, through the application of the procedure introduced in Section 5.5. Based on the analysis of the impact of d^v on the positioning error carried out in Section 6.2.3, the procedure was applied for $d^v = d^v_{max}$.

Figure 15 presents $\beta(d^r, d^v_{max}, \alpha)$ as a function of α for d^r as in TABLE 3. Results show that α between 0.01 and 0.05 leads to values of $\beta(d^r, d^v_{max}, \alpha)$ consistently close to its minimum, for all d^r values and in both SPinV and TWIST testbeds; this suggests that $k_{est} = 0.05 (N^r + N^v)$ is a suitable approximation of k_{opt}. The finding is supported by results shown in Figure 16, comparing the average positioning error $\bar{\epsilon}(d^r, d^v, k_{est})$ vs. $\bar{\epsilon}(d^r, d^v, k_{opt})$ in the SPinV vs. TWIST testbeds. The small difference between the average positioning errors corroborates the reliability of the
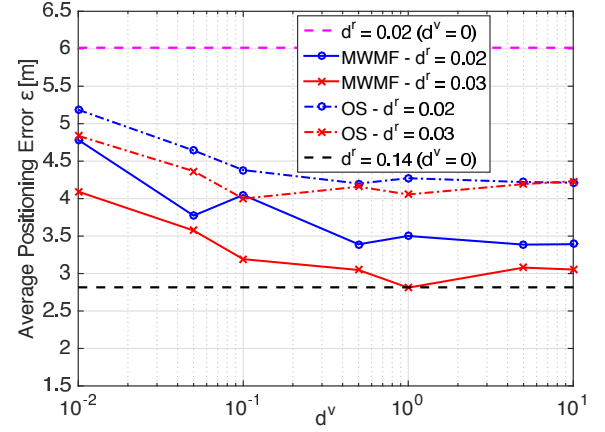

Fig. 13: Average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ as a function of d^v for MWMF vs. OS models in SPinV, for $d^r = 0.02$ and $d^r = 0.03$; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d^r_{min}$ vs. $d^r = d^r_{max}$ are also shown.
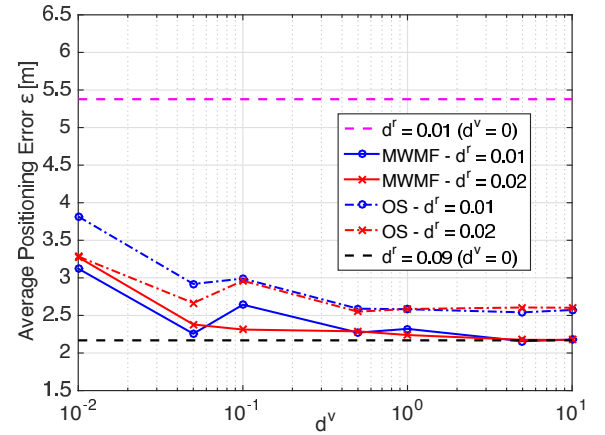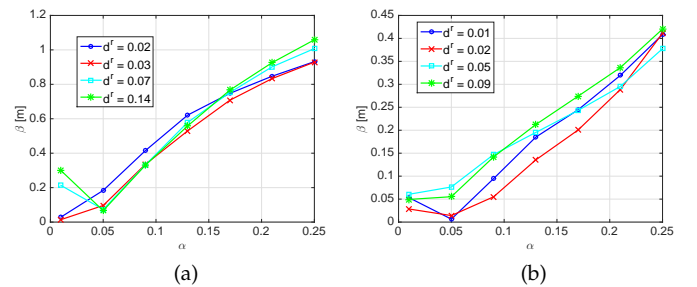

Fig. 14: Average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ as a function of d^v for MWMF vs. OS models in TWIST, for $d^r = 0.02$ and $d^r = 0.03$; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d^r_{min}$ vs. $d^r = d^r_{max}$ are also shown.

(a)

(b)

Fig. 15: $\beta(d^r, d^v_{max}, \alpha)$ as a function of α, for d^r as in TABLE 3 in SPinV (a) vs. TWIST (b).

(a)   (b)

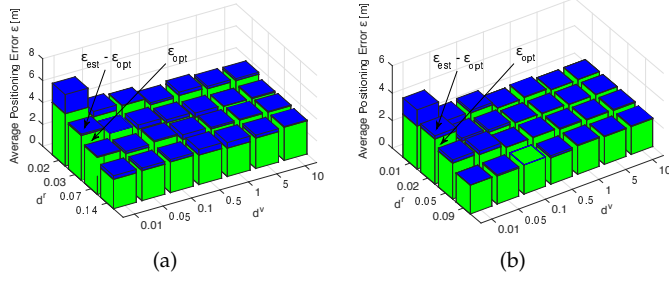Fig. 16: Comparison of average positioning errors $\bar{\epsilon}(d^r, d^v, k_{opt})$ vs. $\bar{\epsilon}(d^r, d^v, k_{est})$ as a function of d^r and d^v, for $\alpha = 0.05$, in SPinV (a) vs. TWIST (b).

proposed k_{opt} estimator and its applicability, with α set to 0.05, to different environments.

6.3 Crowdsourcing-like Scenario

Figure 17 reports the average prediction error $\bar{\delta}(\rho)$ for the SPinV testbed in the Crowdsourcing-like scenario. Comparison against Figure 5, reporting $\bar{\delta}(\rho)$ for the Controlled scenario, highlights that the use of less stable measurements and the introduction of heterogeneity, inherent to crowdsourcing, lead to a degradation in prediction accuracy, particularly relevant for low ρ values and Strategy I.
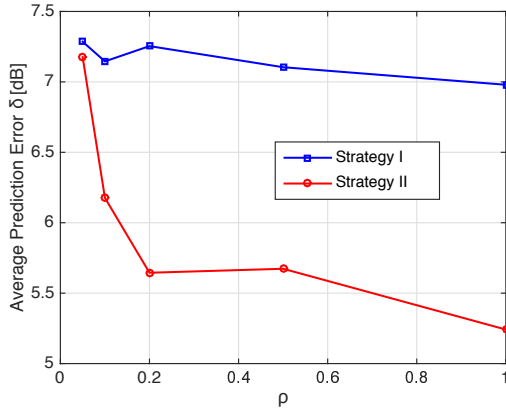

Fig. 17: Average prediction error $\bar{\delta}(\rho)$ as a function of ρ for selection Strategy I vs. II in the Crowdsourcing-like scenario.

The same holds true for the average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ presented in Figure 18. Comparison against Figure 13 confirms that the use of crowdsourced measurements decreases the positioning accuracy, with an increase of the error lower bound $\bar{\epsilon}(d^r_{max}, 0, k_{opt})$ of about 90 cm. ViFi, however, still reaches the new lower bound, confirming its capability to compensate the lack of an exhaustive offline measurement phase with the introduction of virtual fingerprints. The virtualization gain $\mathcal{G}(d^r, d^v, k_{opt})$, reported in Figure 19a, shows similar trends with respect to the Controlled scenario seen in Figure 12, with higher virtualization gains for low d^r, high d^v values.

Finally, Figure 19b compares, in analogy with Figure 16, the average positioning errors $\bar{\epsilon}(d^r, d^v, k_{opt})$ and $\bar{\epsilon}(d^r, d^v, k_{est})$, with k_{est} evaluated with $\alpha = 0.05$: results confirm that the selection of 5% of the total amount of RPs

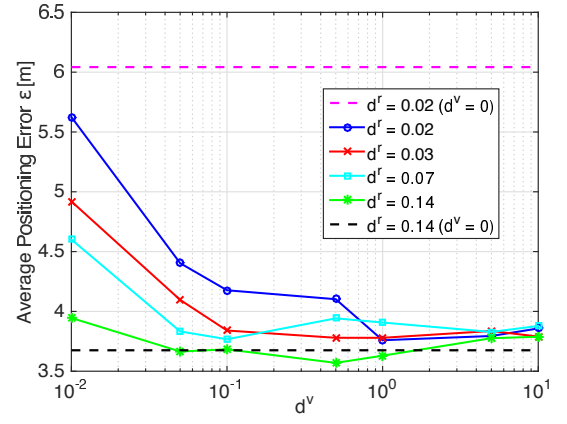Fig. 18: Average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ as a function of d^v, for d^r as in TABLE 3, in the Crowdsourcing-like scenario; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d^r_{min}$ vs. $d^r = d^r_{max}$ are also shown.

(a)   (b)

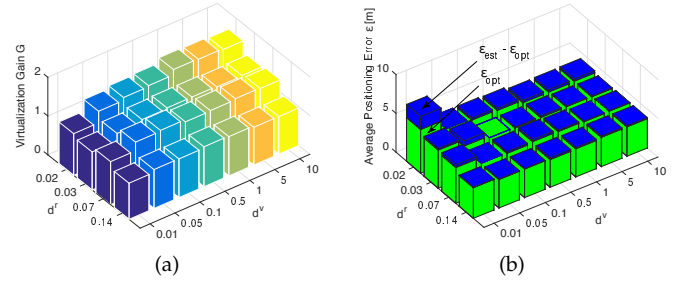Fig. 19: Virtualization gain $\mathcal{G}(d^r, d^v, k)$ as a function of d^r and d^v, for $k = k_{opt}$ (a), and comparison of average positioning error $\bar{\epsilon}(d^r, d^v, k_{opt})$ and $\bar{\epsilon}(d^r, d^v, k_{est})$ for $\alpha = 0.05$ (b), as a function of d^r, d^v in the Crowdsourcing-like scenario.

as value of k leads to an average positioning error within 5% of the positioning error obtained for k_{opt}. In light of the above results, it is possible to conclude that ViFi is applicable to both controlled and crowdsourcing scenarios, and that crowdsourcing might prove a valuable source of data to further reduce the efforts required for the offline phase. Given the significant losses in terms of positioning accuracy caused by the introduction of unreliable measurements in the database, further studies are however required to determine the full extent of the impact of crowdsourced data on virtual fingerprinting systems. An issue particularly relevant for crowdsourcing, but also for fingerprinting in general, is device heterogeneity. Under this respect ViFi does not differ from traditional systems, and solutions recently proposed to address the issue in traditional fingerprinting systems, such as using the difference between RSS values so to erase biases introduced by specific hardware [43], or removing from each RSS value the average RSS received from all APs at the same location [44] might be applied to ViFi as well.

The low number of actual measurements to be taken in the case of ViFi enables however a new solution, unfeasible in traditional fingerprinting, that is collecting multiple sets of measurements with different devices during the

offline phase, and using the best fitting database in the online phase depending on the target device (with device brand/model/chipset derived from the packets exchanged during the association phase).

6.4 Summary of key experimental results

Key results of the analysis presented in the previous subsections can be summarized as follows:

- positioning error decreases as d^v increases, but values of $d^v > 10$ RPs/m^2 do not provide any significant additional improvement (Figures 10 and 11);
- positioning error decreases as d^r increases, but values of $d^r > 0.03$ RPs/m^2 do not provide any additional improvement *if virtual fingerprinting is used* (Figures 13 and 14);
- k_{opt} is directly proportional to $d^r + d^v$, supporting the validity of Equation (10) (Figures 10b and 11b), and $\alpha = 0.05$ consistently leads to a good approximation of k_{opt} across different values of d^r and d^v (Figure 15).

Such results led to the implementation guidelines detailed in Section 7.

7 Implementation Guidelines

The consistency in the behavior of the ViFi system across two different environments suggests the following set of implementation guidelines in a generic environment of area $|\mathcal{A}|$:

1) A density of real RPs of $d^r = 0.03$ RPs/m^2 is sufficient to obtain a reliable RSS prediction using the MWMF model, and the corresponding generation of virtual RPs. Repeated measurements for each RP should be collected and averaged, in order to mitigate the impact of unstable measurements on positioning error.
2) The average RSS prediction error $\bar{\delta}$ is a reliable indicator for choosing the selection strategy leading to the minimization of the average positioning error $\bar{\epsilon}$.
3) A density of virtual RPs of 10 RPs/m^2 is sufficient to achieve a positioning accuracy comparable with a real fingerprinting system operating in the same environment with $d^r = d^r_{max}$.
4) The selection of $k_{est} = \lceil 0.05(d^r + d^v)|\mathcal{A}| \rceil$ in the WkNN online estimator guarantees an average positioning error approximately within 5% of the minimum error achieved with $k = k_{opt}$.

The above guidelines were heuristically derived, and their validity cannot be proved for all testbeds and all environments; however, the following observations support the claim that the guidelines can be applied across different testbeds:

1) The two testbeds do not share any hardware or software component, barring the possibility that the results are depending on a specific combination of infrastructure, devices, or software used during data collection.
2) The two testbeds are deployed in different environments, with different topological characteristics, and different number and positions of APs. These differences are clearly reflected by the different accuracy achievable in the two testbeds, with TWIST consistently

leading to better accuracy compared to SPinV: nonetheless, the above guidelines, when applied to the two testbeds, lead to an average positioning error very close (within 5.5%) to the achievable minimum.
3) The guidelines were applied to a database containing data acquired in a third testbed, the WiLabt [28], and the resulting accuracy is well within 5% of the optimal one, without carrying out any testing phase.

8 ViFi vs. traditional fingerprinting

Information requirements - ViFi requires two additional information pieces when compared to traditional fingerprinting: 1) the position of the APs and 2) the transmit power of the APs. As of 1), the adoption of Simultaneous Location And Mapping (SLAM) techniques might be considered in scenarios where the position of the APs is not known in advance, with a possible loss in accuracy [46]. Regarding 2), the transmit power may be included in the set $\{\mathcal{S}\}$ in the optimization procedure defined by Equation (4), but Strategy II (Specific AP fitting) should be adopted in this case, unless one assumes that W_{TX}^{EIRP}, although unknown, is the same for all APs.

Positioning accuracy - The positioning accuracy in WiFi fingerprinting highly depends on the characteristics of the environment, as shown in [28], [47]. Although an average positioning error of roughly 1 m was achieved in an indoor area of roughly 10x10 m^2, comprised of two rooms and a hallway [45], positioning error increases with an increase in the size of the environment: in large and complex indoor environments, such as those considered in this work, an average positioning error in the order of 2 m is in line with current state of the art. Indeed, in the TWIST testbed similar positioning errors have been demonstrated for various types of localization solutions, including multiple WiFi fingerprinting-based ones [48]; a positioning error above 2 meters was also recently reported in complex environments in [31].

Computational complexity - The complexity of the offline phase in ViFi is determined by the least square fitting procedure (Equation (4)), and by the generation of virtual RPs (Equations (1)-(3)). Following [49], the single fitting procedure of Strategy I has complexity $O(|\mathcal{S}|(N^r L)^2)$, while the L different fitting procedures of Strategy II have complexity $O(|\mathcal{S}|(N^r)^2)$ each. As for the generation of virtual RPs, given the total number $N^v L$ of RSS values to be predicted, and denoted with N_{2D}^{MAX} and N_f^{MAX} the maximum number of 2D objects and floors obstructing the $N^v L$ links, the complexity of the virtual RPs generation is $O(N_{2D}^{MAX} N^v L)$ in terms of multiplications (assuming $N_{2D}^{MAX} > N_f^{MAX}$). The complexity of the offline phase in ViFi is, of course, larger than in a traditional fingerprinting system, where no algorithm is required to fill the database. It should be noted, however, that the offline phase is only carried out at system setup and, if needed, when major changes in the environment are observed. As a result, the increase in complexity can be arguably considered a small price to pay in exchange for the dramatic reduction in time and efforts required for RSS data collection guaranteed by ViFi. Regarding the online phase, according to the analysis presented in [50], the complexity of the WkNN algorithm is de-

termined by a) the computation of the Euclidean distances between the target RSS fingerprint and the RPs (NL multiplications), b) the selection of the k nearest neighbors, (Nk comparisons), and c) the application of the weights to the k neighbors (k multiplications). Since in this work a linear dependency on N is proposed for k, the online complexity is $O(NL)$ multiplications and $O(N^2)$ comparisons.

9 Conclusion

In this work a virtual fingerprinting indoor positioning system, referred to as ViFi, has been proposed. ViFi uses the empirical MWMF indoor propagation model for the generation of virtual RPs and a deterministic Euclidean WkNN algorithm to infer the target location.

The performance of ViFi was experimentally evaluated in two different testbeds, and compared with previous proposals in the literature. Results show that ViFi dramatically reduces the efforts required to setup a WiFi fingerprinting system, without significantly affecting positioning accuracy; in particular, ViFi outperforms virtual fingerprinting systems using simpler propagation models.

Finally, a set of guidelines for the selection of offline and online ViFi parameters was proposed, that saves the additional efforts related to the testing phase typically required for the tuning of a WiFi fingerprinting system. As a result, ViFi can provide the same accuracy of a real fingerprinting system while guaranteeing up to a sevenfold reduction in time and efforts required for measurements.

This work opens the way for research on several topics either already under investigation or being considered for future work. Among them it is worth mentioning: a) adoption of other online estimation algorithms in place of the WkNN algorithm, providing better accuracy and/or lower complexity, and verification of the guidelines in Section 7; b) improved propagation modeling, taking into account device heterogeneity and orientation, particularly relevant in a crowdsourced scenario; c) derivation of analytic models for the design and performance analysis of WiFi fingerprinting systems.

References

[1] Markets and Markets, Indoor Location Market by Component (Technology, Software Tools, and Services), Application, End User (Transportation, Hospitality, Entertainment, Shopping, and Public Buildings), and Region - Global Forecast to 2021, Oct. 2016.

[2] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," in *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[3] V. Honkavirta, T. Perälä, S. Ali-Löytty and R. Piché, "Comparative Survey of WLAN Location Fingerprinting Methods," in *Proc. of IEEE WPNC*, pp. 243–251, Mar. 2009.

[4] M. Kotaru, K. Joshi, D. Bharadia and S. Katti, "SpotFi: Decimeter Level Localization Using WiFi," in *Proc. of ACM SIGCOMM*, pp. 269–282, Aug. 2015.

[5] D. Vasisht, S. Kumar and D. Katabi, "Decimeter-Level Localization with a Single WiFi Access Point," in *Proc. of Usenix NSDI*, pp. 165–178, Mar. 2016.

[6] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based User Location and Tracking System," in *Proc. of IEEE INFOCOM*, vol. 2, pp. 775–784, Mar. 2000.

[7] M. Kessel and M. Werner, "SMARTPOS: Accurate and Precise Indoor Positioning on Mobile Phones," in *Proc. of IARIA MOBILITY*, pp. 158–163, Oct. 2011.

[8] B. Li, J. Salter, A. G. Dempster and C. Rizos, "Indoor Positioning Techniques based on Wireless LAN," in *Proc. of IEEE AusWireless*, pp. 13–16, Mar. 2006

[9] F. Yu, M. Jiang, J. Liang, X. Qin, M. Hu, T. Peng and X. Hu, "5G WiFi Signal-Based Indoor Localization System Using Cluster-Nearest Neighbor Algorithm," in *Hindawi Int'l J. Distrib. Sens. Netw.*, vol. 2014, Article ID 247525, 12 pages, Dec. 2014.

[10] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," in *Springer Int'l J. Wireless Inf. Networks*, vol. 9, no. 3, pp. 155–164, July 2002.

[11] M. A. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. of ACM MobiSys*, pp. 205–218, Jun. 2005.

[12] N. Le Dortz, F. Gain and P. Zetterberg, "WiFi Fingerprint Indoor Positioning System using Probability Distribution Comparison," in *Proc. of IEEE ICASSP*, pp. 2301–2304, Mar. 2012.

[13] M. J. Du and J. Hua, "The design of RFID dual frequency integrated technology based on prison application," in *Proc. of IEEE ITNEC*, pp. 860–863, May 2016.

[14] A. Mathisen, S. Krogh Sorensen, A. Stisen, H. Blunck and K. Gronbaek, "A comparative analysis of Indoor WiFi Positioning at a large building complex," in *Proc. of IEEE/GRSS IPIN*, pp. 1–8, Oct. 2016.

[15] R. Yasmine and L. Pei, "Indoor fingerprinting algorithm for room level accuracy with dynamic database," in *Proc. of IEEE/GNSS UPINLBS*, pp. 113–121, Nov. 2016.

[16] P. Bolliger, "Redpin - adaptive, zero-configuration indoor localization through user collaboration," in *Proc. of ACM MELT*, pp. 55–60, Sept. 2008.

[17] A. Rai, K. K. Chintalapudi, V. N.. Padmanabhan and R. Sen. "Zee: zero-effort crowdsourcing for indoor localization," in *Proc. of ACM MobiCom*, pp. 293–304, Aug. 2012.

[18] G. Chatzimilioudis, A. Konstantinidis, C. Laoudias and D. Zeinalipour-Yazti, "Crowdsourcing with Smartphones," in *IEEE Internet Comput.*, vol. 16, no. 5, pp. 36–44, Sept.-Oct. 2012.

[19] S. Yang, P. Dessai, M. Verma and M. Gerla, "FreeLoc: Calibration-free crowdsourced indoor localization," in *Proc. of IEEE INFOCOM*, pp. 2481–2489, Apr. 2013.

[20] C. Laoudias, D. Zeinalipour-Yazti and C. G. Panayiotou, "Crowd-sourced indoor localization for diverse devices through radiomap fusion," in *Proc. of IEEE/GRSS IPIN*, pp. 1–7, Oct. 2013.

[21] G. Caso and L. De Nardis, "Virtual and Oriented WiFi Fingerprinting Indoor Positioning based on Multi-Wall Multi-Floor Propagation Models," to appear in *Springer Mob. Netw. Appl. (MONET)*, SM-168, 2017.

[22] Widyawan, M. Klepal, D. Pesch, "Influence of Predicted and Measured Fingerprint on the Accuracy of RSSI-based Indoor Location Systems," in *Proc. of IEEE WPNC*, pp. 145–151, Mar. 2007.

[23] E. Damosso, L. M. Correia *et al.*, "COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report," European Commission, 1999.

[24] A. Borrelli, C. Monti, M. Vari and F. Mazzenga, "Channel models for IEEE 802.11b indoor system design," in *Proc. of IEEE ICC*, pp. 3701–3705, June 2004.

[25] G. Caso and L. De Nardis, "On the Applicability of Multi-wall Multi-floor Propagation Models to WiFi Fingerprinting Indoor Positioning," in *Proc. of EAI FABULOUS*, LNICST vol. 159, pp. 166–172, Sept. 2015.

[26] F. Lemic, J. Büsch, M. Chwalisz, V. Handziski and A. Wolisz, "Infrastructure for Benchmarking RF-based Indoor Localization under Controlled Interference," in *Proc. of IEEE UPINLBS*, pp. 26–35, Nov. 2014.

[27] T. Van Haute, E. De Poorter, F. Lemic, V. Handziski, N. Wirstöm, T. Voigt, A. Wolisz and I. Moerman, "Platform for benchmarking of RF-based indoor localization solutions," in *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 126–133, Sept. 2015.

[28] F. Lemic, V. Handziski, G. Caso, P.. Crombez, L. De Nardis, A. Wolisz, T. Van Heute and E. De Poorter, "Toward Extrapolation of WiFi Fingerprinting Performance Across Environments," in *Proc. of AMC HotMobile*, pp. 69–74, Feb. 2016.

[29] A. K. M. M. Hossain, H. N. Van, Y. Jin, W. S. Soh, "Indoor Localization Using Multiple Wireless Technologies," in *Proc. of IEEE MASS*, pp. 1–8. Oct. 2007.

[30] N. Hernández, M. Ocaña, J. M. Alonso and E. Kim, "WiFi-based Indoor Localization Using a Continuous Space Estimator From Topological Information," in *Proc. of IEEE/GRSS IPIN*, pp. 148899-1-148899-4, Oct. 2015.

[31] S. Kumar, R Hegde, and N. Trigoni, "Gaussian Process Regression for Fingerprinting based Localization," in *Elsevier Ad Hoc Netw.*, vol. 51, pp. 1–10, Nov. 2016.

[32] Y. Gwon and R. Jain, "Error characteristics and calibration-free techniques for wireless LAN-based location estimation," in *Proc. of ACM MOBIWAC*, pp. 2–9, Oct. 2004.

[33] T. Pulkkinen, J. Verwijnen and P. Nurmi, "WiFi positioning with propagation-based calibration," in *Proc. of ACM IPSN*, pp. 366–367, Apr. 2015.

[34] S. Kumar and R. Hegde, "An Efficient Compartmental Model for Real-Time Node Tracking Over Cognitive Wireless Sensor Networks," in *IEEE Trans. Signal Process.*, vol.63, no.7, pp.1712–1725, Sept. 2015.

[35] K. Chintalapudi, A. P. Iyer, V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. of ACM MobiCom*, pp. 173–184, Sept. 2010.

[36] H. Lim, L. C. Kung, J. C. Hou and H. Luo, "Zero-Configuration, Robust Indoor Localization: Theory and Experimentation," in *Proc. of IEEE INFOCOM*, pp. 1–12, Apr. 2006.

[37] Y. Ji, S. Biaz, S. Pandey and P. Agrawal, "ARIADNE: a dynamic indoor signal map construction and localization system," in *Proc. of ACM MobiSys*, pp. 151–164, June 2006.

[38] C. Wu, Z. Yang, Y. Liu and W. Xi, "WILL: Wireless Indoor Localization without Site Survey," in *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 839–848, Apr. 2013.

[39] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte and J. Huerta, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," in *Elsevier Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, Dec. 2015.

[40] G. Caso, L. De Nardis and M.-G. Di Benedetto, "A Mixed Approach to Similarity Metric Selection in Affinity Propagation-Based WiFi Fingerprinting Indoor Positioning," in *MDPI Sensors*, vol. 15, no. 11, pp. 27692–27720, Oct. 2015.

[41] G. Caso, L. De Nardis and M.-G. Di Benedetto, "Frequentist inference for WiFi fingerprinting 3D indoor positioning," in *Proc. of IEEE ICC Workshops*, pp. 809–814, June 2015.

[42] O. Baala, Y. Zheng and A. Caminada, "The Impact of AP Placement in WLAN-Based Indoor Positioning System," in *Proc. of IEEE ICN*, pp. 12–17, Mar. 2009.

[43] A. K. M. M. Hossain, Y. Jin, W.-S. Soh, and H. N. Van, "SSD: a Robust RF Location Fingerprint Addressing Mobile Devices' Heterogeneity," in *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 65–77, Nov. 2011.

[44] C.-H. Wang, T.-W. Kao, S.-H. Fang., Y. Tsao, L.-C. Kuo, K. Shih-Wei and N.-C. Lin, "Robust Wi-Fi location fingerprinting against device diversity based on spatial mean normalization," in *Proc. of IEEE APSIPA*, pp. 1–4, Jan. 2013.

[45] D. Lymberopoulos, et al., "A Realistic Evaluation and Comparison of Indoor Location Technologies: Experiences and Lessons Learned," in *Proc. of IEEE/ACM IPSN*, pp. 178–189, Apr. 2015.

[46] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii." in *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108-117, Sept. 2006.

[47] T. Van Haute, E. De Poorter, I. Moerman, F. Lemic, V. Handziski, A. Wolisz, N. Wirstrom and T. Voigt, "Comparability of RF-based Indoor Localization Solutions in Heterogeneous Environments: An Experimental Study," in *Inderscience Int'l J. Ad Hoc Ubiquitous Comput.*, vol. 23, no. 1/2, pp. 92–114, Jan. 2016.

[48] F. Lemic, V. Handziski, A. Wolisz, T. Constambeys, C. Laoudias, S. Adler, S. Schmitt and Y. Yang, "Experimental Evaluation of RF-based Indoor Localization Algorithms Under RF Interference," in *Proc. of IEEE ICL-GNSS*, pp. 1–8, June 2015.

[49] G. Simon, R. Pintelon, L. Sujbert and J. Schoukens, "An Efficient Nonlinear Least Square Multisine Fitting Algorithm," in *IEEE Trans. Instrum. Meas.*, vol. 51, no. 4, pp. 750–755, Aug. 2002.

[50] W. Zuo, D. Zhang and K. Wang, "On kernel difference-weighted k-nearest neighbor classification," in *Springer Pattern Anal. Applic.*, vol. 11, no. 3/4, pp. 247-257, Sept. 2008.

Giuseppe Caso received the M.Sc and Ph.D. degrees from Sapienza University of Rome, Italy, in 2012 and 2016, respectively. He is a Post-Doctoral Fellow at the DIET Department of Sapienza University of Rome. From 2012 to 2015, he has held visiting positions at Leibniz University of Hannover, Germany, King's College London, UK, and Technical University of Berlin, Germany. His research interests include indoor positioning technologies and methodologies, with particular focus on WiFi and UWB fingerprinting, indoor channel modeling, UWB communications, cognitive radio and context-aware networking. He is an IEEE Member.

Luca De Nardis (M'98) is an Assistant Professor with the DIET Department, Sapienza University of Rome. He received the Laurea and Ph.D. degrees from the Sapienza University of Rome, Italy, in 2001 and 2005, respectively. In 2007, he was a Post-Doctoral Fellow with the EECS Department, University of California at Berkeley, USA. He authored or co-authored over 100 publications in international peer-reviewed journals and conferences. His research interests focus on UWB radio technology, medium access control, routing and positioning protocols for wireless networks, and on the design of cognitive wireless networks.

Filip Lemic is a research assistant and a Ph.D. candidate at the Telecommunication Networks Group of Technical University of Berlin, Germany. He received the B.Sc. and M.Sc. degrees from the University of Zagreb, Croatia. In 2015 and 2016, he was a visiting research assistant at the Berkeley Wireless Research Center and the Qualcomm SWARMLab at the University of California at Berkeley, USA. His research interests include location and context awareness, Internet of Things, and mmWave communication. He is a member of IEEE and ACM.

Vlado Handziski received the M.Sc. degree from Ss. Cyril and Methodius University in Skopje, Republic of Macedonia, and Ph.D. degree from Technical University of Berlin, Germany, in 2002 and 2011, respectively. He is a Senior Researcher in the Telecommunication Networks Group at Technical University of Berlin, Germany, where he coordinates the activities in the areas of sensor networks, cyber-physical systems, and the Internet of Things. He is also serving as Interim Professor at the chair for Embedded Systems at Technical University of Dresden, Germany. He is an IEEE Member.

Adam Wolisz received his degrees (Diploma 1972, Ph.D. 1976, Habil. 1983) from Silesian University of Technology, Gliwice, Poland. He joined Technical University of Berlin, Germany in 1993, where he is a Chaired Professor in telecommunication networks and Executive Director of the Institute for Telecommunication Systems. He is also an Adjunct Professor at the Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA. His research interests are in architectures and protocols of communication networks. He is an IEEE Senior Member.

Maria-Gabriella Di Benedetto obtained her Ph.D. degree in 1987 from Sapienza University of Rome, Italy. In 1991, she joined the Faculty of Engineering of Sapienza University of Rome, where she is a Full Professor of telecommunications. She has held visiting positions at the Massachusetts Institute of Technology, USA, the University of California, Berkeley, USA, and the University of Paris XI, France. In 1994, she received the Mac Kay Professorship award from the University of California, Berkeley. Her research interests include wireless communication systems, in particular impulse radio communications, and speech. She is an IEEE Fellow.