

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 multiple independent 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 other proposals, that may achieve submeter accuracy but require specific hardware/software modifications to devices [4], [5]. WiFi fingerprinting traditionally operates in two phases. During the so-called *offline phase*, RSS values (*fingerprints*) from WiFi Access Points (APs) detected in the environment 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 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 cardinality and positions of the RPs, and number of measurements collected at each RP [6], [7].

Regarding the online phase, deterministic and probabilistic k -Nearest Neighbors (k NN) and Weighted k NN (Wk NN) algorithms are by far the most widely proposed and investigated: on the 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, characterized by restricted access areas, where measurements cannot be collected;

- 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: giuseppe.caso@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 –.

emergency and diagnostic rooms in hospitals, where activities cannot be interrupted by measurement collection; **large shopping malls and skyscrapers** spanning over tens or even hundreds 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 either eliminating or reducing the offline phase were proposed.

As regards the elimination of the offline phase, most of the works in the literature proposed strategies to model a direct relation between RSS values and spatial coordinates, in order to infer the distances between APs and target device, and in turn target position. Simple radio propagation modeling and regression [16], [17], along with either modifications of off-the-shelf devices [16], or additional hardware [18], [19], [20], were proposed. However, the increase of positioning error [16], [17], and the need for enhanced system components [18], [19], [20], challenge the adoption of such approaches in place of traditional fingerprinting.

Considering the reduction of the offline phase, two main solutions emerged:

Crowdsourcing – In this approach, devices that make use of the IPS contribute to the collection of location-dependent RSS samples [21], in either a voluntary or involuntary way [22]. Application of crowdsourcing entails further challenges, in terms of uncertainty of location of RPs and corresponding RSS values, caused by lack of control on the mobile device, and heterogeneity of devices used for collection [23], [24], [25].

Virtual Fingerprinting – In this approach, the generation of so-called *virtual* RPs¹ is carried out in the area of interest, by adopting RSS prediction methods based on either *propagation modeling*, in which an empirical radio propagation model is trained with an initial set of measurements, or by *interpolation*, in which adjacent *real* RPs are interpolated. In both cases, virtual RPs, along with real RPs, whose amount is hopefully smaller when compared to traditional fingerprinting, are then jointly used in the online phase.

In this paper, a WiFi IPS based on *virtual fingerprinting by propagation modeling*, referred to as ViFi, is proposed and analyzed. The system adopts RSS propagation modeling for the offline phase reduction, and combines it with a deterministic position estimation algorithm.

Design strategies adopted for ViFi were principally driven by the following considerations:

- 1) Differently from offline phase elimination schemes discussed above, virtual fingerprinting does not directly infer distances. This in turn allows the use of advanced propagation models, embedded with multiple propagation and topological parameters, that may lead to improved prediction of RSS values, accurate generation of virtual RPs, and thus enhanced positioning accuracy.
- 2) In contrast to offline phase reduction schemes by either crowdsourcing or virtual fingerprinting by interpola-

tion, virtual fingerprinting by propagation modeling allows the deployment of IPSs in scenarios where an extensive and spatially regular data collection during the offline phase is prohibitive or even impossible. RSS prediction by advanced propagation models can be in fact obtained even in areas where no measurements are collected, if measurements in environments with similar characteristics, e.g. different floors in the same building, are available, as analyzed in [26], [27].

Overall, ViFi offline and online phases were designed aiming to limit the complexity and avoid device modifications and additional hardware. In particular, 1) the Multi-Wall Multi-Floor (MWMF) was selected among several empirical models, since it provides advanced propagation modeling at a reasonable degree of complexity [26], [28], [29], [30], [31], and 2) a traditional, low-complex Wk NN algorithm was adopted for the fingerprinting online phase.

ViFi was implemented and tested in two testbeds with different AP topology and signal coverage, 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. The validity of the proposed guidelines was then corroborated by deploying ViFi in a third testbed, with results presented in [32].

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, complexity, and vulnerability, while Section 9 concludes the paper and highlights possible future research lines.

2 Virtual fingerprinting by RSS prediction

2.1 RSS Prediction by Interpolation

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 [33]. Experimental results showed, in both cases, a positioning error decrease when compared with fingerprinting systems using a low amount of real RPs. Moreover, virtual fingerprinting via Support Vector Regression (SVR) and Gaussian Processes (GP) was recently proposed in [34] and [35], respectively. In the first case, a slightly better accuracy with respect to a traditional k NN estimator was obtained; in the second case, a 30% accuracy improvement, with respect to the Horus system [11], was achieved. As already discussed, a spatially regular collection of real RPs is in general required for the application of interpolation techniques.

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.

2.2 RSS Prediction by Indoor Propagation Modeling

An empirical propagation model taking into account the effect of obstructing walls on the perceived RSS was preliminary 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) 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 [28] the impact of using 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 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 [36], where a log-distance model was adopted, limiting the optimization to the path loss exponent. A more complex model was proposed in [37], 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.

2.3 Open issues and proposed contribution

The focal issue emerging from the analysis of previous work is related to the difficulty in achieving a satisfactory trade off between complexity and accuracy of fingerprinting systems. On the one hand, as observed in Section 1, the elimination of the offline phase either leads to accuracy decrease or requires enhanced components, while crowdsourcing introduces further challenges. On the other hand, virtual fingerprinting may provide reasonable solutions to significantly reduce the offline phase, but strong experimental evidence is still missing. As a matter of fact, given a propagation model or an interpolation technique, the impact of number and positions of a) real RPs on the RSS prediction accuracy, and b) virtual RPs on the positioning accuracy is still unclear.

As a second issue, it can be observed that, in general, no guidelines are provided for setting up a virtual fingerprinting system in different environments, given that further investigations are required in order to uncover the relation between offline and online system parameters. This limitation leads to a case by case choice rather than systematic settings,

e.g. in terms of number, position and spatial distribution of collected real RPs, generated virtual RPs, and the value to assign to k in case a k NN/ Wk NN estimator is adopted.

A third issue regards the need for sets of measurements acquired in a controlled and reliable fashion, common to all proposals reviewed in Sections 2.1 and 2.2. This is in contrast with recent proposals that allow for data collection in an uncontrolled fashion, such as crowdsourcing. The joint use of crowdsourcing with virtual fingerprinting is indeed all but unexplored in the literature, and its impact on RSS prediction and positioning accuracy is unknown.

As regards the first issue, the contribution of this paper is the definition of a novel virtual fingerprinting system, ViFi, that provides an accuracy comparable to real fingerprinting systems based on exhaustive RSS collections, while significantly reducing the amount of data to be collected. This is obtained by adopting the MWMF model to generate virtual RPs, that has a complexity lower than the techniques in [34], [35], and comparable to the models proposed in [6], [36], while leading to better RSS prediction. With respect to [8], [28], [33], this paper proposes an in-depth experimental analysis on both adopted propagation model and proposed system, focusing on the impact of densities and positions of real and virtual RPs on RSS prediction and positioning accuracy.

As regards the second issue, the relation between system parameters is analyzed, and an empirical and effective rule for the selection of the value of k to be used in the Wk NN estimator, adopted in ViFi for the online phase, is derived and experimentally validated. Altogether, thanks to an extensive experimental analysis, this paper proposes a full set of guidelines for the implementation and setup of ViFi in different environments, indicating: a) the required 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. With respect to the last issue, this paper provides preliminary results on the combination of virtual fingerprinting and crowdsourcing techniques for offline phase reduction, focusing in particular on the aspect of uncertainty of RSS values, inherent to crowdsourcing, and thus offering a framework for further research activities.

3 Multi-Wall Multi-Floor Propagation Models

Multi-Wall Multi-Floor models [29] 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 [30]. MWMF models take into account objects, that obstruct signal propagation in an indoor wireless link, leading to the following path loss model [29]:

$$PL_{MWMF} = PL_{OS} + A_{MWMF} \quad [\text{dB}], \quad (1)$$

where the PL_{OS} term models the path loss over the Tx-Rx distance d , while A_{MWMF} models the additional loss due to obstructing obstacles. The PL_{OS} term is defined according to the One Slope (OS) model, as follows [29]:

$$PL_{OS}(d, \gamma) = l_0 + 10\gamma \log(d) \quad [\text{dB}], \quad (2)$$

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 [30]). When, in the most general case, Tx and Rx are located on different floors, A_{MWMF} is given by [29]:

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

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 MWMF requires an initial set of M measurements, and for each measurement ($m = 1, 2, \dots, M$) 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:

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

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) $W_{\text{TX}}^{\text{EIRP}}$ at distance d_m . $\hat{\text{RSS}}_m$ is computed as follows:

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

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 \mathbf{s}_{n_1} is associated with the n_1 -th RP. The generic \mathbf{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 accuracy vs. time and effort devoted to measurements. Since only a subset of the L APs may be detected at the generic RP, the \mathbf{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 $\hat{\mathbf{s}}_{n_2}$ contains the predicted RSS at the n_2 -th virtual RP from the l -th AP.

4.2 Online Phase

ViFi adopts a deterministic $Wk\text{NN}$ estimator using combination of real vs. virtual RPs, to infer target location.

Denoting by $N = N^r + N^v$ the number of RPs in the area \mathcal{A} , \mathbf{s}_n ($n = 1, 2, \dots, N$) the RSS fingerprint of n -th RP, and \mathbf{s}_i the RSS fingerprint collected during the i -th positioning request by a target device in unknown position $\mathbf{p}_i = (x_i, y_i, z_i)$, position estimation relies on the computation of a similarity metric $\text{sim}_{n,i} = \text{sim}(\mathbf{s}_n, \mathbf{s}_i)$. The $Wk\text{NN}$ algorithm selects the k RPs that present the highest $\text{sim}_{n,i}$ values and provides an estimate of \mathbf{p}_i defined as:

$$\hat{\mathbf{p}}_i = \frac{\sum_{n=1}^k (\text{sim}_{n,i}) \mathbf{p}_n}{\sum_{n=1}^k \text{sim}_{n,i}}, \quad (6)$$

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

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

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

Typical orders are $o = 1$ (inverse Manhattan distance) and $o = 2$ (inverse Euclidean distance, used in ViFi). Similarity metrics using modified versions of the inner product between RSS vectors have also been proposed [38], [39].

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:

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

In [26], [31], several strategies were proposed for deriving the MWMF propagation parameters; among them, the following two strategies were adopted in this paper:

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

Strategy II: Specific AP Fitting – The set of measurements from a specific AP to all RPs is 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 is defined as follows:

$$d^v = \frac{N^v}{|\mathcal{A}|}. \quad (9)$$

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 grid scheme widely adopted for real RPs, and random placement.

4.4 Online Phase Implementation Options

The $WkNN$ algorithm foresees the selection of two parameters: 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 [39], [40], 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 as k_{opt} . As shown in [3], [6], [7], [8], [38], [40], 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 $k_{opt} = f(d^r, d^v)$. In this work a linear law is proposed for $f(\cdot, \cdot)$, leading to an approximate k_{opt} value, k_{est} , as follows:

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

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 RSS variation 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 d^r ; second, virtual fingerprinting only makes sense if d^r is much smaller than d^v , 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 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 the adopted testbeds.

5 Experimental Analysis Setup

This section describes the SPinV and TWIST testbeds, implemented at Sapienza University of Rome and Technical University of Berlin, respectively. These two environments were adopted in the initial testing of the ViFi system². In the present section, the procedures defined for the experimental analysis of ViFi are also 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 m² each. $L_1 = 6$ APs working @ 2.4 and 5 GHz, and $L_2 = 7$ 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 1-st and 2-nd floor, respectively. In this work, the SPinV 2-nd floor has been adopted as evaluation area \mathcal{A} , and the APs on this floor have been considered in the fingerprinting measurement campaign, so that $L = L_2$. Two different measurement campaigns were carried out, each corresponding to a different application scenario:

Controlled scenario – In this scenario, during the offline phase, $N^{t,tot} = 72$ RPs were selected for fingerprints collection on a regular grid within \mathcal{A} ; fingerprints were also collected in a set of $N^t = 31$ Test Points (TPs) randomly distributed over \mathcal{A} . All measurements were carried out during weekend afternoons, in order to mitigate the effect of radio interference due to possible mobile and temporary connection points. No particular mitigation strategy was instead applied as regards the interference with other APs available in the area. Weekend campaigns also allowed to minimize the variation of propagation conditions due to human presence, considering that no other person than the single human surveyor was present during the RSS collection. A MacBook Pro equipped with an AirPort Extreme Network Interface Card was used, placed on a wooden platform 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. The RSS collection required approximately 6 minutes for each RP location.

² Description of the w-iLab.t I testbed [27], [41], implemented at iMinds in Ghent, Belgium, and used to further assess the ViFi performance, is reported in [32], together with the obtained results.

Crowdsourcing-like scenario – In this scenario, $N^{r,\text{tot}} = 69$ RPs and $N^t = 26$ TPs were collected in the area³. All measurements were performed during weekdays, by a single human surveyor equipped with an Android-based Samsung tablet. In this case, no radio and human interference mitigation strategies were adopted, in order to mimic a realistic crowdsourcing scenario. Normal human office activity (several static and walking persons randomly distributed in the environment) was observed during the RSS collection. RPs fingerprints were obtained by averaging $q = 5$ measurements, while TPs fingerprints include a single measurement. A specific characteristic of the SPinV testbed is the location of APs. On both floors, APs are 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 [41], [42]. TWIST is deployed in an office environment and covers one floor with an area of about $30 \times 15 \text{ m}^2$. Within this environment, $L = 4$ dedicated WiFi were placed at known positions and configured to operate @ 2.4 GHz with a beacon transmission period of $T_b = 100 \text{ ms}$ and a transmit power $W_{\text{TX}}^{\text{EIRP}} = 20 \text{ dBm}$. 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, $N^{r,\text{tot}} = 41$ RPs were identified on a regular grid and measured within \mathcal{A} ; an additional set of $N^t = 80$ uniformly distributed TPs was collected.

In the case of TWIST, all measurements were taken with the same settings defined for the Controlled scenario in SPinV, and no Crowdsourcing-like scenario was considered⁴.

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

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

In agreement with the approach proposed in [26], [31], 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) of Section 3, out of the total number of RPs, $N^{r,\text{tot}}$, so that $N^r = \lceil \rho N^{r,\text{tot}} \rceil$. The 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 $N^{r,\text{tot}}$ RPs, of the

3. Crowdsourced RPs/TPs have the same positions of Controlled scenario RPs/TPs; a few measurements are missing due to practical constraints encountered during the collection campaigns.

4. RSS data collected in SPinV and TWIST testbeds, as well as in w-iLab.t analyzed in [32], are available for download at http://newyork.ing.uniroma1.it/ViFi_-_Supplemental_Material_-_Datasets.zip.

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.

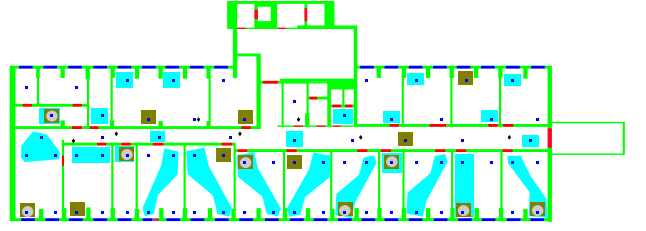


Fig. 1: Position of real RPs (blue squares) and APs (black diamonds) in the SPinV testbed. Areas of different colors highlight selected RPs when $\rho = 0.1$ (light grey), $\rho = 0.2$ (dark green), $\rho = 0.5$ (light blue), $\rho = 1$ (white + all colors).

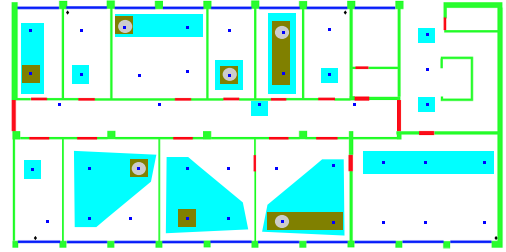


Fig. 2: Position of real RPs (blue squares) and APs (black diamonds) in the TWIST testbed. Areas of different colors highlight selected RPs when $\rho = 0.1$ (light grey), $\rho = 0.2$ (dark green), $\rho = 0.5$ (light blue), $\rho = 1$ (white + all colors).

- 2) The model fitting procedure defined by Equation (4), 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 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 $N^{r,\text{tot}}$ RPs were originally collected. Next, the accuracy of the model was evaluated by defining the prediction error as follows:

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

where $\hat{s}_{l,n}(\rho)$ is the predicted RSS for the generic (AP $_l$, RP $_n$) pair, obtained by using a set of $N^r = \lceil \rho N^{r,\text{tot}} \rceil$

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 l -th AP, the cumulative distribution function (CDF) of $\delta_l(\rho)$ and the average error $\bar{\delta}_l(\rho) = \frac{\sum_{n=1}^{N_r^{\text{tot}}} \delta_{l,n}(\rho)}{N_r^{\text{tot}}}$ were also evaluated.

Results of the analysis are reported 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 ($i = 1, 2, \dots, N^t$) 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 $(x_i, y_i, z_i) = \mathbf{p}_i$ and $(\hat{x}_i, \hat{y}_i, \hat{z}_i) = \hat{\mathbf{p}}_i$ are the actual and the estimated positions of the i -th target device, respectively. Note that dependence of $\hat{\mathbf{p}}_i$ 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 $\bar{\epsilon}(d^r, d^v, k) = \frac{\sum_{i=1}^{N^t} \epsilon_i(d^r, d^v, k)}{N^t}$ were 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 reported 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:

- 1) Given a generic (d^r, d^v) combination, k_{opt} and in turn $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ were evaluated;
- 2) For each $\alpha \in [\alpha_{\min} : \alpha_{\max}]$, $k_{\text{est}}(\alpha)$ was evaluated following Equation (10), and the corresponding average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{est}}(\alpha))$ was determined;
- 3) 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.

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}
$\{\alpha_{\min}, \alpha_{\max}\}$	{0.01, 0.25}	

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 model proposed in [28], [36].

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 $\delta_l(\rho)$ values with respect to Strategy II.

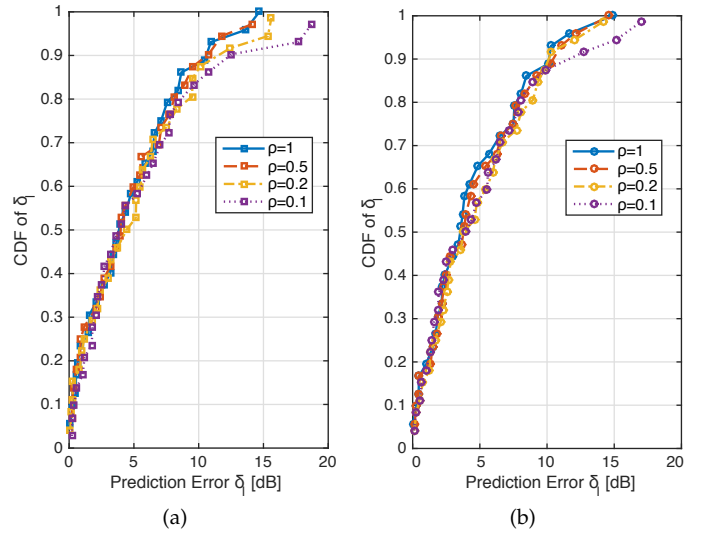


Fig. 3: Cumulative Distribution Function of prediction error $\delta_l(\rho)$ for a reference AP in SPinV. Strategy I (a) vs. Strategy II (b).

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 $\delta_l(\rho)$ 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. 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 far

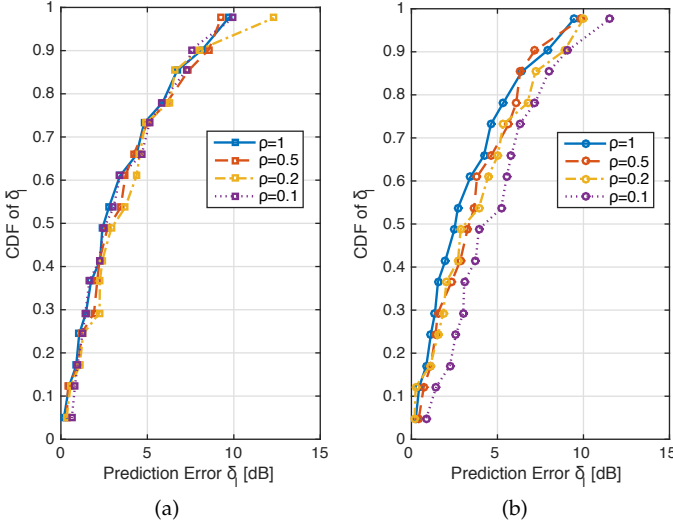


Fig. 4: Cumulative Distribution Function of prediction error $\delta_l(\rho)$ for a reference AP in TWIST. Strategy I (a) vs. Strategy II (b).

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 $\delta_l(\rho)$ for a reference AP and $\rho = 1$ for MWMF vs. OS models. 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

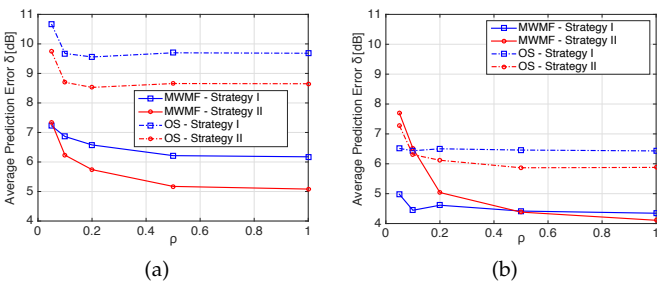


Fig. 5: Average prediction error $\bar{\delta}(\rho)$ as a function of ρ for Strategies I and II in SPinV (a) vs. TWIST (b) (Continuous lines: proposed MWMF model; dashed-dotted lines: OS model proposed in [28], [36].)

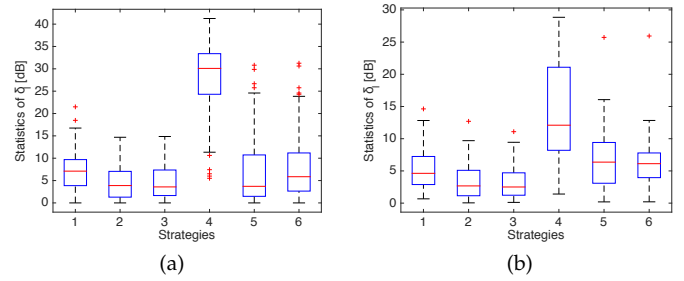


Fig. 6: Statistics of $\delta_l(\rho)$ for a reference AP and $\rho = 1$ in SPinV (a) vs. TWIST (b) (1: MWMF-No Fit; 2: MWMF-Strategy I; 3: MWMF-Strategy II; 4: OS-No Fit; 5: OS-Strategy I; 6: OS-Strategy II).

predictions were obtained in MWMF by using propagation parameters estimated for a totally different area \mathcal{A}' , reported in [30]; in the OS model the settings $\gamma = 2$, $l_0 = 20$ were adopted. Note that in Figure 6, and following ones, error statistics are summarized in the form of boxplot diagrams showing minimum, maximum and median values, 25-th and 75-th percentiles, 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 models.

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 decreases 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 [43].

Figures 7b and 8b present the boxplot of the positioning error $\epsilon(d^r, 0, k_{opt})$ as a function of d^r , and confirm that in real fingerprinting systems the value of k_{opt} spans in the [2, 10] range, as observed in [3], [6], [7], [8], [38], [40].

6.2.2 Impact of δ^r of real RPs selection strategies and virtual RPs placement

Figure 9 reports the average positioning error $\bar{\epsilon}(d^r, d_{max}^r, k_{opt})$ for Strategy I and II, and d^r as in TABLE

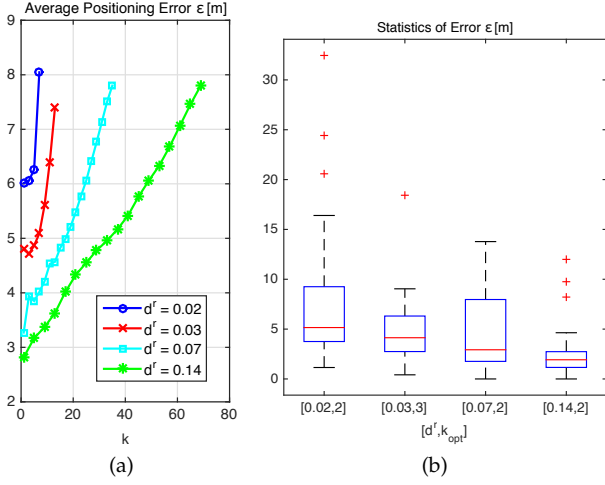


Fig. 7: Average positioning error $\bar{\epsilon}(d^r, 0, k)$ as a function of k , for d^r as in TABLE 3 (a) and statistics of $\epsilon(d^r, 0, k_{\text{opt}})$ (b) in SPinV.

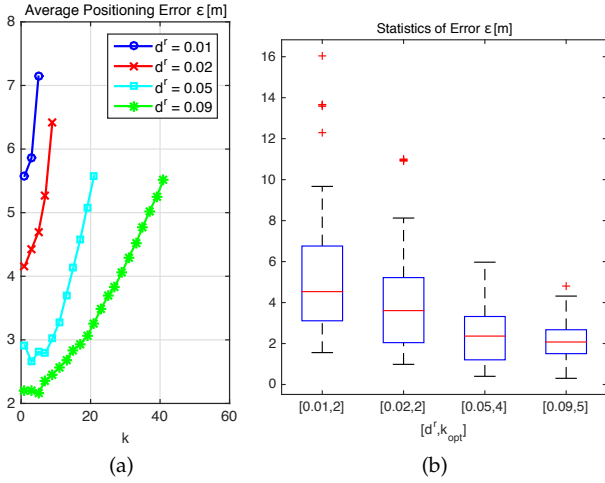


Fig. 8: Average positioning error $\bar{\epsilon}(d^r, 0, k)$ as a function of k , for d^r as in TABLE 3 (a), and statistics of $\epsilon(d^r, 0, k_{\text{opt}})$ (b) in TWIST.

3. Results are in agreement with Figure 5: within the SPinV environment, Strategy II outperforms Strategy I, and strategies performance approach when $d^r = d^r_{\text{min}}$; on the contrary, within the TWIST environment, Strategy I leads to better positioning accuracy, and strategies performance approach when $d^r = d^r_{\text{max}}$. 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.

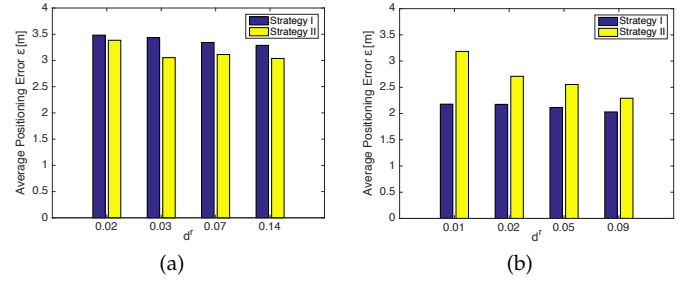


Fig. 9: Average positioning error $\bar{\epsilon}(d^r, d^v_{\text{max}}, k_{\text{opt}})$, for d^r as in TABLE 3, and for selection Strategy I and II, in SPinV (a) vs. TWIST (b) (Blue bars: Strategy I; yellow bars: Strategy II).

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 $\bar{\epsilon}(d^r_{\text{min}}, d^v, k)$ as a function of k (Figures 10a and 11a) and the boxplot of the positioning error $\epsilon(d^r_{\text{min}}, d^v, k_{\text{opt}})$ (Figures 10b 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 in Figure 12, showing the virtualization gain $\mathcal{G}(d^r, d^v, k_{\text{opt}})$, as defined in Equation (13), for both testbeds, and confirm that the introduction of virtual RPs significantly improves positioning accuracy for low d^r 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).

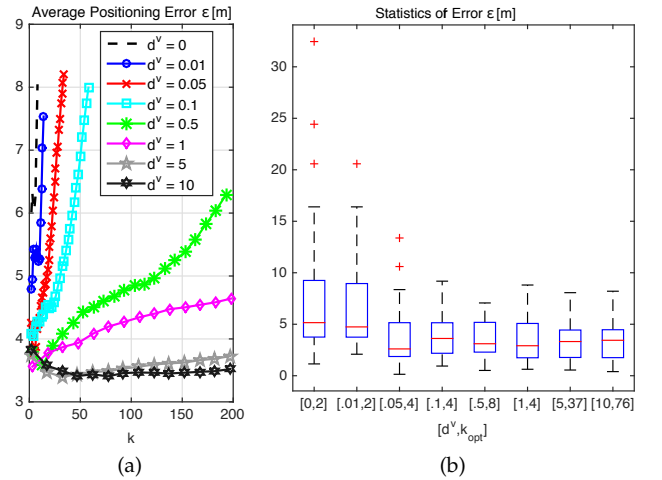


Fig. 10: Average positioning error $\bar{\epsilon}(d^r_{\text{min}}, d^v, k)$ as a function of k , for d^v as in TABLE 3 (a), and statistics of $\epsilon(d^r_{\text{min}}, d^v, k_{\text{opt}})$ (b) in SPinV.

Low d^r values were thus selected for further investigation, with results presented in Figure 13 and Figure 14, showing the average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ 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 [28], [36], as well as for a real fingerprinting system, with $d^r = d^r_{\text{min}}$ and $d^r = d^r_{\text{max}}$ as upper and lower bounds for positioning error, respectively.

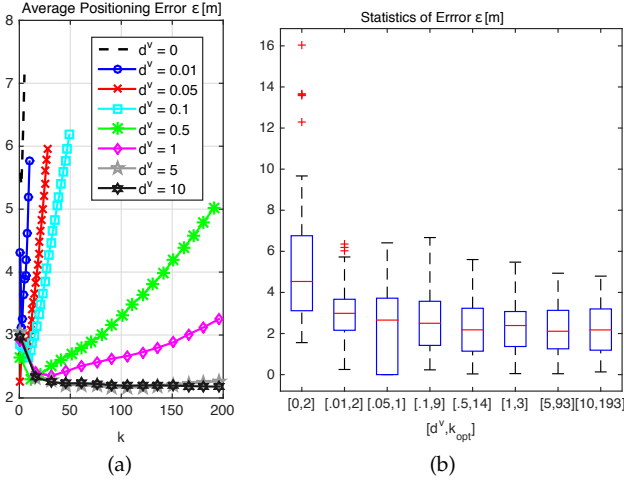


Fig. 11: Average positioning error $\bar{\epsilon}(d^r_{\min}, d^V, k)$ as a function of k , for d^V as in TABLE 3 (a), and statistics of $\epsilon(d^r_{\min}, d^V, k_{opt})$ (b) in TWIST.

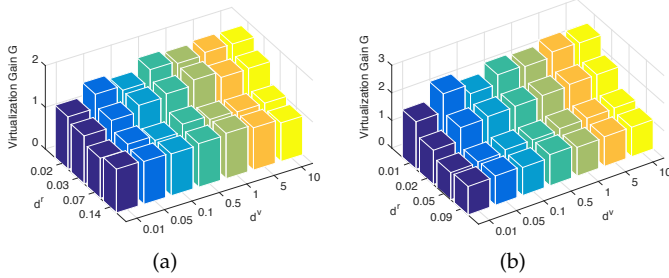


Fig. 12: Virtualization gain $\mathcal{G}(d^r, d^V, k)$ as a function of d^r and d^V , for $k = k_{opt}$, in SPinV (a) vs. TWIST (b).

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 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 1) 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 2) 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 of 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

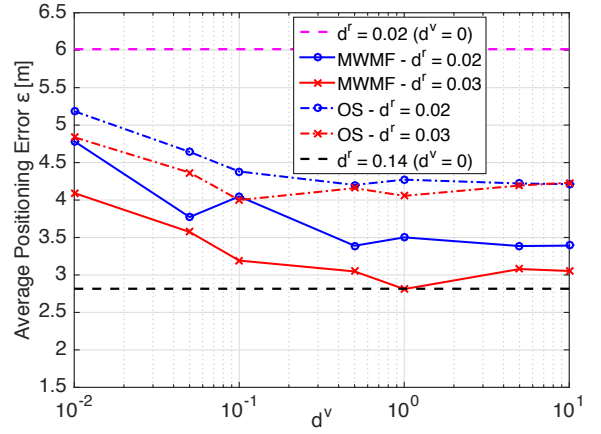


Fig. 13: Average positioning error $\bar{\epsilon}(d^r, d^V, k_{opt})$ as a function of d^V for MWMF vs. OS models in SPinV, for $d^r = 0.02$ and $d^r = 0.03$; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d^r_{\min}$ vs. $d^r = d^r_{\max}$ are also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

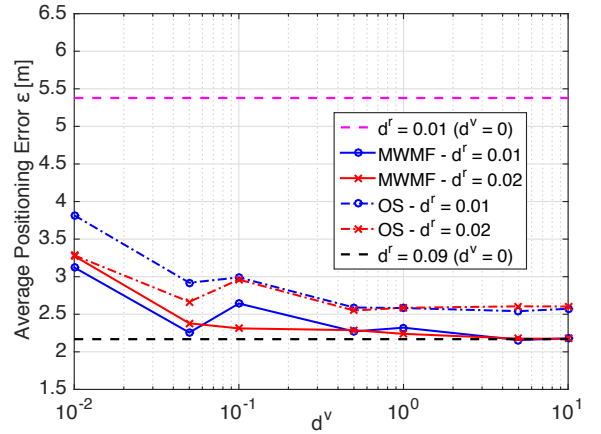


Fig. 14: Average positioning error $\bar{\epsilon}(d^r, d^V, k_{opt})$ as a function of d^V for MWMF vs. OS models in TWIST, for $d^r = 0.01$ and $d^r = 0.02$; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d^r_{\min}$ vs. $d^r = d^r_{\max}$ are also shown.

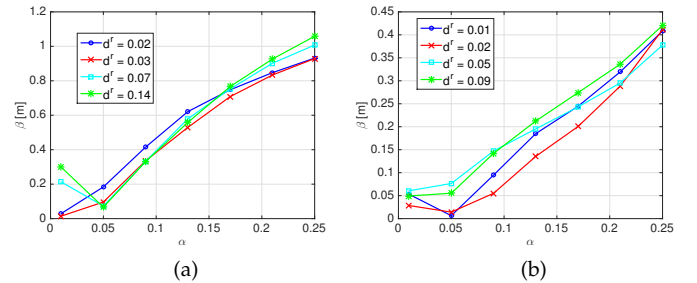


Fig. 15: $\beta(d^r, d^V_{\max}, \alpha)$ as a function of α , for d^r as in TABLE 3 in SPinV (a) vs. TWIST (b).

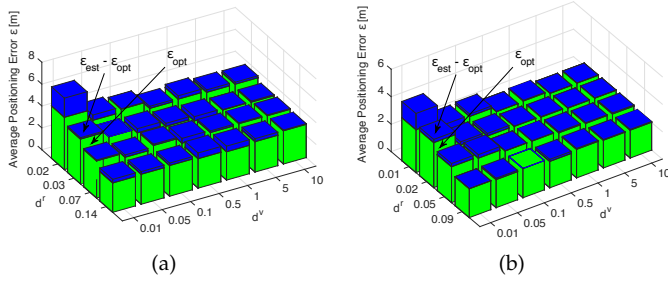


Fig. 16: Comparison of average positioning errors $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ vs. $\bar{\epsilon}(d^r, d^v, k_{\text{est}})$ as a function of d^r and d^v , for $\alpha = 0.05$, in SPinV (a) vs. TWIST (b).

TWIST testbeds; this suggests that $k_{\text{est}} = 0.05(N^r + N^v)$ is a suitable approximation of k_{opt} . The finding is supported by results shown in Figure 16, comparing the average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{est}})$ vs. $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ in the SPinV vs. TWIST testbeds. The small difference between the average positioning errors corroborates the reliability of the 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 MWMF average prediction error $\bar{\delta}(\rho)$ for the SPinV testbed in the Crowdsourcing-like scenario, in comparison with the error obtained in the Controlled scenario, already presented in Figure 5. Figure highlights that the use of less stable measurements, inherent to crowdsourcing, leads to a degradation in prediction accuracy.

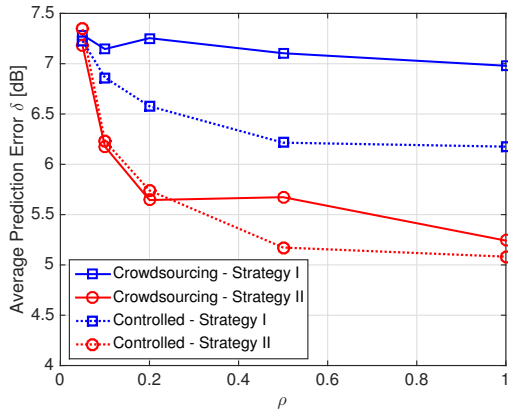


Fig. 17: Average prediction error $\bar{\delta}(\rho)$ as a function of ρ for selection Strategy I vs. II in the Crowdsourcing-like scenario.

The same holds true for the average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ 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 $\bar{\epsilon}(d_{\text{max}}^r, 0, k_{\text{opt}})$ 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 $\mathcal{G}(d^r, d^v, k_{\text{opt}})$, 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.

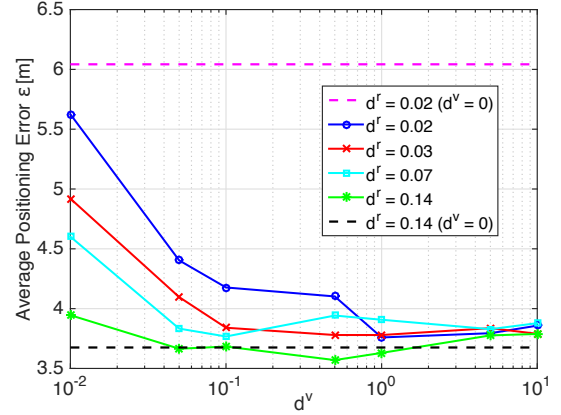


Fig. 18: Average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ as a function of d^v , for d^r as in TABLE 3, in the Crowdsourcing-like scenario; upper and lower bounds on positioning error observed for a real fingerprinting system with $d^r = d_{\text{min}}^r$ vs. $d^r = d_{\text{max}}^r$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

Finally, Figure 19b compares, in analogy with Figure 16, the average positioning errors $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ and $\bar{\epsilon}(d^r, d^v, k_{\text{est}})$, with k_{est} evaluated with $\alpha = 0.05$: results confirm that the selection of 5% of the total amount of RPs 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 option to further reduce the offline phase. Given the significant losses in terms of positioning accuracy caused by the introduction of unreliable measurements, further studies are however required to determine the full extent of the impact of crowdsourced data on virtual fingerprinting systems.

6.4 Summary of key experimental results

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

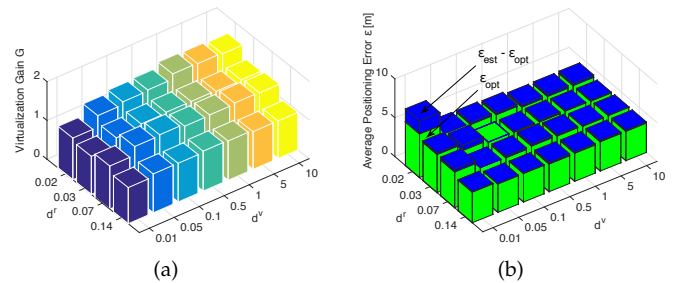


Fig. 19: Virtualization gain $\mathcal{G}(d^r, d^v, k)$ as a function of d^r and d^v , for $k = k_{\text{opt}}$ (a), and comparison of average positioning error $\bar{\epsilon}(d^r, d^v, k_{\text{opt}})$ and $\bar{\epsilon}(d^r, d^v, k_{\text{est}})$ for $\alpha = 0.05$ (b), as a function of d^r, d^v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

- positioning error decreases as d^r increases, but values of $d^r > 10$ RPs/m² do not provide significant additional improvement (Figures 10 and 11);
- positioning error decreases as d^r increases, but values of $d^r > 0.03$ RPs/m² do not provide 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 16).

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

7 Implementation Guidelines

The consistency in the ViFi behavior 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/m² 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 from 1 to 10 RPs/m² is sufficient to achieve a positioning accuracy comparable with a real fingerprinting system operating in the same environment with $d^r = d^r_{\text{max}}$.
- 4) The selection of $k_{\text{est}} = \lceil 0.05(d^r + d^v)|\mathcal{A}| \rceil$ in the Wk NN estimator guarantees an average positioning error within 5% of the error achieved with $k = k_{\text{opt}}$.

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 of applicability across different testbeds:

- 1) SPinV and TWIST 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) SPinV and TWIST 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) As mentioned above, the guidelines were applied to a third testbed, w-iLab.t I, and experimental results fully corroborate the proposed guidelines [32].

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 [44]. Regarding 2), the transmit power may be included in the set $\{\mathcal{S}\}$ in the optimization procedure defined by Equation (4), but Strategy II is preferable in this case, unless one assumes that $W_{\text{TX}}^{\text{ELRP}}$, although unknown, is the same for all APs.

Positioning accuracy – The positioning accuracy in WiFi fingerprinting depends on the characteristics of the environment, as shown in [27], [45]. Although an average positioning error of about 1 m was achieved in an indoor area of roughly 10×10 m² [46], positioning error increases as the environment size increases: 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, similar positioning errors have been recently demonstrated for several localization solutions in TWIST [47] and other environments [35].

Complexity and Scalability – 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 [48], 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 $O(N_{2D}^{\text{MAX}} N^v L)$ in terms of multiplications (assuming $N_{2D}^{\text{MAX}} > N_f^{\text{MAX}}$). The complexity of the ViFi offline phase is, by definition, larger than in traditional fingerprinting, where no computations are required to fill the database. This increase in complexity is however largely compensated by the dramatic reduction in time and efforts required for RSS collection. As a reference, the offline phase of ViFi in the SPinV testbed would require about 45 minutes instead of about 7 hours. As a conclusion, ViFi outperforms traditional fingerprinting in terms of system implementation scalability. A better scalability enables an efficient deployment of ViFi in scenarios where traditional fingerprinting would be unpractical due to the size of the area or to the need for frequent radiomap updates to cope with a variable environment.

Regarding the online phase, according to the analysis presented in [49], the complexity of the Wk NN algorithm is determined 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. The complexity of the ViFi online phase is exactly the same as of traditional fingerprinting adopting a Wk NN algorithm, but ViFi may require a slightly longer execution time due to the larger set of real + virtual RPs. However, several optimization mechanisms can be easily applied to ViFi, aiming to improve its usage scalability. For example, in so-called *two-step* Wk NN schemes [39], the complexity can be significantly reduced by performing a clustering of RPs via Affinity Propagation algorithm, and adopting *coarse*

and *fine* positioning steps for the online phase.

Vulnerability – A plethora of works investigate the vulnerabilities of fingerprinting systems [2]. Two vulnerabilities particularly relevant to WiFi can be identified: environment variations over time and device heterogeneity. *Environmental changes over time* may impact positioning accuracy by invalidating the RSS data, requiring thus a radiomap update. In this respect, WiFi provides two major improvements. First, if the environment change is restricted to a specific topological variation (e.g. a single wall being removed), the underlying MWMF model allows to update the WiFi radiomap by simply regenerating the set of virtual RPs using the new topology, without the need for a new channel training. Second, even if major changes occur, calling for a new channel training, WiFi requires a significantly lower number of real RPs to operate, thus reducing the time and effort required to collect data and update the radiomap, so to restore the system to its best performance. *Measurement mismatch due to device heterogeneity* may also negatively impact accuracy. Solutions proposed to address this issue in real fingerprinting systems, such as using the difference between RSS values so to erase biases introduced by specific hardware [50], or removing from each RSS value the average RSS received from all APs at the same location [51] might be easily applied to WiFi as well. Moreover, the low number of measurements to be collected in WiFi enables a new solution, unpractical in traditional fingerprinting: multiple sets of measurements with different devices can be collected during the offline phase while still saving significant time and effort with respect to the extensive data collection campaign of a real fingerprinting system. In the online phase, the database that best fits the target device can be then adopted to mitigate the impact of device heterogeneity.

9 Conclusion

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

The performance of WiFi was experimentally evaluated in multiple independent testbeds, and compared with previous proposals in the literature. Results show that WiFi outperforms virtual fingerprinting systems using simpler propagation models, and provides the same accuracy of a real fingerprinting system while guaranteeing up to a sevenfold reduction in time and efforts for measurements.

A set of guidelines for the selection of offline and online WiFi parameters was also proposed, that saves the additional efforts related to the testing phase typically required for tuning a WiFi fingerprinting system.

This work opens the way for further research on several topics, including: a) adoption of other online estimation algorithms in place of the Wk NN 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 enhanced models for the design and 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] S. He and S.-H G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," in *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, Firstquarter 2016.
- [3] V. Honkavirta *et al.*, "Comparative Survey of WLAN Location Fingerprinting Methods," in *Proc. of IEEE WPNC*, pp. 243–251, Mar. 2009.
- [4] M. Kotaru *et al.*, "SpotFi: Decimeter Level Localization Using WiFi," in *Proc. of ACM SIGCOMM*, pp. 269–282, Aug. 2015.
- [5] D. Vasishth, 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 *et al.*, "Indoor Positioning Techniques based on Wireless LAN," in *Proc. of IEEE AusWireless*, pp. 13–16, Mar. 2006.
- [9] F. Yu *et al.*, "5G WiFi Signal-Based Indoor Localization System Using Cluster-Nearest Neighbor Algorithm," in *Int'l J. Distrib. Sens. Netw.*, vol. 2014, Article ID 247525, 12 pages, Dec. 2014.
- [10] T. Roos *et al.*, "A Probabilistic Approach to WLAN User Location Estimation," in *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 *et al.*, "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] 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.
- [17] K. Chintalapudi, A. P. Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. of ACM MobiCom*, pp. 173–184, Sept. 2010.
- [18] H. Lim *et al.*, "Zero-Configuration, Robust Indoor Localization: Theory and Experimentation," in *Proc. of IEEE INFOCOM*, pp. 1–12, Apr. 2006.
- [19] Y. Ji *et al.*, "ARIADNE: a dynamic indoor signal map construction and localization system," in *Proc. of ACM MobiSys*, pp. 151–164, June 2006.
- [20] C. Wu *et al.*, "WILL: Wireless Indoor Localization without Site Survey," in *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 839–848, Apr. 2013.
- [21] P. Bolliger, "Redpin - adaptive, zero-configuration indoor localization through user collaboration," in *Proc. of ACM MELT*, pp. 55–60, Sept. 2008.
- [22] A. Rai *et al.*, "Zee: zero-effort crowdsourcing for indoor localization," in *Proc. of ACM MobiCom*, pp. 293–304, Aug. 2012.
- [23] G. Chatzimilioudis *et al.*, "Crowdsourcing with Smartphones," in *IEEE Internet Comput.*, vol. 16, no. 5, pp. 36–44, Sept.-Oct. 2012.
- [24] S. Yang *et al.*, "FreeLoc: Calibration-free crowdsourced indoor localization," in *Proc. of IEEE INFOCOM*, pp. 2481–2489, Apr. 2013.
- [25] C. Laoudias, D. Zeinalipour-Yazti, and C. G. Panayiotou, "Crowdsourced indoor localization for diverse devices through radiomap fusion," in *Proc. of IEEE/GRSS IPIN*, pp. 1–7, Oct. 2013.
- [26] G. Caso and L. De Nardis, "Virtual and Oriented WiFi Fingerprinting Indoor Positioning based on Multi-Wall Multi-Floor Propagation Models," in *Mob. Netw. Appl. (MONET)*, vol. 22, no. 5, pp. 825–833, 2017.

- [27] F. Lemic *et al.*, "Toward Extrapolation of WiFi Fingerprinting Performance Across Environments," in *Proc. of AMC HotMobile*, pp. 69–74, Feb. 2016.
- [28] Widyawan, M. Klepal, and 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.
- [29] E. Damosso *et al.*, "COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report," European Commission, 1999.
- [30] A. Borrelli *et al.*, "Channel models for IEEE 802.11b indoor system design," in *Proc. of IEEE ICC*, pp. 3701–3705, June 2004.
- [31] 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.
- [32] G. Caso *et al.*, "ViFi: Virtual Fingerprinting WiFi-based Indoor Positioning via Multi-Wall Multi-Floor Propagation Model – Supplemental Material", 2018.
- [33] A. K. M. M. Hossain *et al.*, "Indoor Localization Using Multiple Wireless Technologies," in *Proc. of IEEE MASS*, pp. 1–8, Oct. 2007.
- [34] N. Hernández *et al.*, "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.
- [35] S. Kumar, R. Hegde, and N. Trigoni, "Gaussian Process Regression for Fingerprinting based Localization," in *Ad Hoc Netw.*, vol. 51, pp. 1–10, Nov. 2016.
- [36] T. Pulkkinen, J. Verwijnen, and P. Nurmi, "WiFi positioning with propagation-based calibration," in *Proc. of ACM IPSN*, pp. 366–367, Apr. 2015.
- [37] 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.
- [38] J. Torres-Sospedra *et al.*, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," in *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, Dec. 2015.
- [39] 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 *Sensors*, vol. 15, no. 11, pp. 27692–27720, Oct. 2015.
- [40] 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.
- [41] T. Van Haute *et al.*, "Platform for benchmarking of RF-based indoor localization solutions," in *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 126–133, Sept. 2015.
- [42] F. Lemic *et al.*, "Infrastructure for Benchmarking RF-based Indoor Localization under Controlled Interference," in *Proc. of IEEE UPINLBS*, pp. 26–35, Nov. 2014.
- [43] 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.
- [44] 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.
- [45] T. Van Haute *et al.*, "Comparability of RF-based Indoor Localization Solutions in Heterogeneous Environments: An Experimental Study," in *Int'l J. Ad Hoc Ubiquitous Comput.*, vol. 23, no. 1/2, pp. 92–114, Jan. 2016.
- [46] 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.
- [47] F. Lemic *et al.*, "Experimental Evaluation of RF-based Indoor Localization Algorithms Under RF Interference," in *Proc. of IEEE ICL-GNSS*, pp. 1–8, June 2015.
- [48] G. Simon *et al.*, "An Efficient Nonlinear Least Square Multisine Fitting Algorithm," in *IEEE Trans. Instrum. Meas.*, vol. 51, no. 4, pp. 750–755, Aug. 2002.
- [49] W. Zuo, D. Zhang, and K. Wang, "On kernel difference-weighted k-nearest neighbor classification," in *Pattern Anal. Applic.*, vol. 11, no. 3/4, pp. 247–257, Sept. 2008.
- [50] A. K. M. M. Hossain *et al.*, "SSD: a Robust RF Location Fingerprint Addressing Mobile Devices' Heterogeneity," in *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 65–77, Nov. 2011.

- [51] C.-H. Wang *et al.*, "Robust Wi-Fi location fingerprinting against device diversity based on spatial mean normalization," in *Proc. of IEEE APSIPA*, pp. 1–4, Jan. 2013.



Giuseppe Caso received the M.Sc and Ph.D. degrees from Sapienza University of Rome in 2012 and 2016. He is a Post-Doctoral Fellow at the DIET Department, Sapienza University of Rome. From 2012 to 2015, he has held visiting positions at Leibniz University of Hannover, King's College London, and Technical University of Berlin. His research interests include WiFi and UWB indoor positioning, cognitive radio, and context-aware, distributed communications. 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 Sapienza University of Rome in 2001 and 2005. In 2007, he was a Post-Doctoral Fellow with the EECS Department, University of California, Berkeley. 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.

wireless networks, and



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



Vlado Handziski received the M.Sc. degree from Ss. Cyril and Methodius University in Skopje, and Ph.D. degree from Technical University of Berlin, in 2002 and 2011. He is a Senior Researcher in the Telecommunication Networks Group, Technical University of Berlin, 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. He is an IEEE Member.



Adam Wolisz received his degrees (Diploma 1972, Ph.D. 1976, Habil. 1983) from Silesian University of Technology, Gliwice. He joined Technical University of Berlin 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. His research interests are in architectures and protocols of communication networks. He is an IEEE Senior Member.



communication systems, in particular impulse radio communications, and speech. She is an IEEE Fellow.