

WI-LO: Wireless Indoor LOcalization through Multi-Source Radio Fingerprinting

Marco Di Felice*, Carlos Bocanegra[†] and Kaushik Roy Chowdhury[†]

* Department of Computer Science and Engineering, University of Bologna, Italy

Email: difelice@cs.unibo.it

[†] Department of Electrical and Computer Engineering, Northeastern University, Boston, USA

Email: bocanegra.c@husky.neu.edu, krc@ece.neu.edu

Abstract—The paper answers a fundamental question: can the accuracy of Wi-Fi based localization be significantly increased by fusing information from alternate sources like LTE signals and magnetometers, collected through software defined radios and smart-devices? Further, it aims to eliminate dependency of well established localization techniques on only Wi-Fi access point (AP) positions, and instead, proposes a diversity-leveraging architecture called as the Wireless Locator (Wi-LO) framework. Wi-LO is a client-server paradigm for indoor localization that achieves precision by pattern-matching the collected signal samples with a priori references for each transmitter type, thus giving a stand-alone decision for these diverse sensing modes. The novelty of the paper is fusing these decision outcomes and resolving mis-matches (for e.g., Wi-Fi and LTE suggest different locations) in a seamless manner, and also identifying the best source to use at each location based on its spatial resolution. Wi-Lo is rigorously evaluated on a test-bed, with the proposed scheme of combining Wi-Fi, LTE and magnetometer performing better localization than the classical Wi-Fi-only approach in both urban (+8%) and rural (+25%) scenarios.

I. INTRODUCTION

Location-Aware Services (LAS) based on GPS have become extremely popular thanks to increasing adoption of smart devices. An exciting domain within LAS is identifying virtual boundaries, so called geofencing within indoor environments. From business (e.g., proximity advertisements and in-store analytics) to emergency communications (e.g., building evacuation), several scenarios may leverage indoor LAS. As a result, many techniques for indoor localization has been proposed, highly heterogeneous in terms of technologies, complexity and cost, as summarized in [1].

Recently, Wi-Fi fingerprinting techniques based on the received signal strength (RSS) have gained considerable attention, justified by the lack of reliable GPS indoors but also by the high density of access points (APs). Radio fingerprinting algorithms typically work in offline or online phases [2] [3]. During the offline phase, a radio-frequency (RF) map of the target environment is built by sensing the Wi-Fi channels at pre-decided reference points (RPs). During the online phase, the RSS signal is sampled at the unknown location, and then compared with the stored data in order to infer the current user location. Well-known localization systems like RADAR [4] and FreeLoc [5] are based on Wi-Fi radio fingerprinting. There have been additional efforts to speed up the offline phase through crowdsensing [6] [7] [8] and using pattern matching

schemes, often employing machine learning (ML) [9] [10]. However, most of these localization techniques are highly dependent on specific topologies and on the location of the APs [11] [12]: a limiting situation on rural or suburban areas that typically have low density of APs.

Our proposed approach is motivated by the capabilities within software defined radios and smart-devices that are equipped with multiple radio interfaces and several embedded sensors, like the accelerometer and the gyroscope. The latter can be used to detect movement and then track the user's trajectory [13] [14]. Multiple radio technologies (e.g., LTE, Wi-Fi and Bluetooth Low Energy or BLE) have been used to create richer RF maps, demonstrated previously for static localization [15] [16]. In this paper, we use this second approach as our starting point, i.e., we study whether radio fingerprinting techniques based on the concurrent utilization of all the radio interfaces typically present in a smart device (WiFi, LTE and BLE), in addition to the magnetometer (MAG) can enhance the localization accuracy in both high- and low-density of APs.

There are three main contributions in this paper:

- First, we investigate the impact of AP density and of the training parameters, such as number of samples and RPs on the overall localization performance. We use extensive simulations over varying network topologies, and then average the results over thousands of trials in order to infer general properties.
- Second, we describe a client-server paradigm called Wireless LOcator (Wi-LO) that extracts radio fingerprints from Wi-Fi, LTE, BLE and MAG signal samples, applies pattern matching individually on each data source, and then merges the results to detect the user location. As opposed to the data-fusion scheme proposed in [17] [21], our scheme measures the signals from each source in isolation, and then combines the hard decisions. As a result, Wi-LO is extensible in the sense that it can be augmented with different fingerprint patterns unique to emerging wireless technologies. In the present implementation, Wi-LO aims at quantifying the signal features at each RP and for each source of samples. This is used to formulate the metrics that can appropriately weight the sources during the fusion process.
- Third, we evaluate the performance of three different pattern matching and three fusion schemes, and we show that the selective weighting and fusion in Wi-LO can enhance the

accuracy of localization compared to pure Wi-Fi in both urban (+8%) and rural (+25%) scenarios.

The rest of the paper is structured as follows. Section II reviews the state of art of indoor localization techniques, focusing on data-fusion approaches. Section III illustrates the performance of Wi-Fi radio fingerprinting by a simulation study. Section IV describes the architecture of the Wi-LO software. Section V evaluates different pattern matching and fusion algorithms on a real test-bed. Conclusions follow in Section VI.

II. RELATED WORKS

Wi-Fi-based fingerprinting is a well-known indoor localization technique, used by RADAR [4] and FreeLoc [5]. Despite the intuitive conceptual idea, its implementation may present several challenges, as surveyed in [2] [3]. One major issue in simplistic extension to other radio interfaces is the dependence of the localization accuracy on specific hardwares and wireless standards, even if they are incorporated in today's commercial smart devices [2].

Current research on radio fingerprinting addresses both the phases of radio map training and matching. [6] and [7] investigate the utilization of crowdsensing for the RF map building in the offline phase. Similarly, [8] proposes automatic update strategies by using static mobile devices. In [11], Welch's t-test is used to detect significant changes in the Wi-Fi links for triggering a map update. The pattern matching algorithm responsible for comparing the stored and sensed radio fingerprints is a key function in the online phase. [9] compares several different machine learning algorithms, and shows that the K^* algorithm gives the best performance in terms of localization accuracy. [10] compares classical Bayesian-based matching schemes with non Bayesian approaches based on the Dempster-Shafer framework.

Data-fusion schemes combine Wi-Fi radio fingerprints with external data provided by the additional hardware available on modern smartphones. We distinguish among three classes of data-fusion approaches: multimedia enhanced, multi-sensors and multi-radio. An example of first class is the WAIPO system [18], which combines Wi-Fi and MAG signals along with a photo-matching feature using the camera of the smart device. Multi-sensor approaches exploit the readings of the accelerometer/gyroscope in order to track the trajectory of mobile users [13] [14]. Our paper belongs to multi-radio approach, as it performs data-fusion of radio fingerprints gathered by heterogeneous wireless technologies. A multi-modal localization system integrating Wi-Fi and digital TV measurements is described in [15]. In [16], the authors show that the positioning error can be decreased by a factor equal to three by using both LTE and Wi-Fi radio fingerprints. In [17], MAG measurements are analyzed in order to create a sparse RF map where field anomalies are identified as location-specific signatures. The most similar works to our paper are [20] and [21], since they combine all the radio interfaces available on smart devices (Wi-Fi, BLE and LTE). In [20], three localization techniques are designed based on a cascade

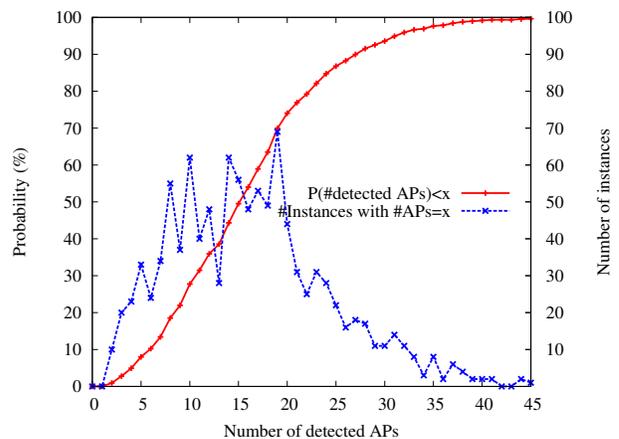


Fig. 1. The CDF of the number of Wi-Fi APs in the downtown area of Bologna, Italy.

approach. In [21] the heterogeneous radio fingerprints are merged by using kernel matrices. The proposed Wi-LO system provides the following novelties compared to the existing literature: (i) it treats each source in isolation, so that it is highly customizable based on smart device hardware characteristics and user's preference; (ii) it considers the different spatial resolution provided by different sources (e.g. LTE and Wi-Fi), while gaining maximum benefit from the presence of identifiable signatures on each RF map.

III. WI-FI ONLY LOCALIZATION

In this Section, we study the performance of Wi-Fi radio fingerprinting techniques and motivate the need of multi-radio data-fusion approaches such as [19]. Most indoor localization schemes are based on Wi-Fi considering the significant density of Wi-Fi APs in highly populated urban environments. Figure 1 shows the Cumulative Distribution Function (CDF) of the number of APs in the city-center of Bologna, Italy, which constitutes a densely inhabited area (around 2400 inhabitants/Km²). We averaged more than 1200 spectrum measurements performed at different locations in the city center. We notice that the number of APs typically exceeds 15 units at each location. While this practical study shows that Wi-Fi AP deployments are widespread, few works have analyzed in depth the impact of AP density on the performance of radio fingerprinting techniques, and the relationship between the parameters used during the offline phase (e.g. number of samples) that impacts localization accuracy. While several test-beds and experimental works have been reported in the literature, these results are strongly tied to the scenario and to the specific location of the APs. Differently from theoretical frameworks like [11] [12], we undertake a simulation study with a large number of topologies. More specifically, we model in NS2 a 3D square urban environment of side equal to P_{eside} meters¹. At the center, we place a target square building

¹We use the notation P_{var} to indicate the parameter var which can be tuned during the online/offline phase.

of P_{nfloor} floors and of side equal to P_{tside} . Each floor of the target building is divided into square rooms of side equal to P_{rside} . We then consider a random deployments of P_{nap} APs over the scenario. The APs can be located within the target building or within any other neighboring buildings. The software developed for this simulation study allows modeling both the offline phase (i.e. map construction) and online phase (i.e. user detection). During the offline phase, each floor of the target building is divided into a grid of square cells, denoted as Reference Points (RPs). Let P_{rpside} be the cell size, i.e. the RP granularity. Moreover, let P_{rp} be the set of RPs in the target building. Thus, we get: $|P_{rp}| = \lceil (\frac{P_{tside}}{P_{rpside}}) \rceil^2 \cdot P_{nfloor}$. At each RP i , P_{off} samples of the Wi-Fi spectrum are obtained. Let $|AP_i^{off}|$ be the set of APs detected at RP i . For each AP $j \in AP_i^{off}$, we store the following information:

$$R_j^{off} = \langle BSSID_j^{off}, RSS_{ij}^{off}, \mu_{ij}^{off}, \sigma_{ij}^{off} \rangle \quad (1)$$

where $BSSID_j^{off}$ is the MAC address of AP j , RSS_{ij}^{off} is the set of RSS samples, μ_{ij}^{off} and σ_{ij}^{off} are respectively the mean and variance values over RSS_{ij}^{off} . The RSS_{ij}^{off} values are obtained by extending the path loss model in [22], and by taking into account the attenuation caused by the distance, by indoor/outdoor walls, by internal floors and the correlated shadowing at different locations, modeled as a Gaussian variable with zero mean. Let P_σ denote the maximum shadowing variance over all the RPs.

During the online phase, we randomly generate the user position u within the target building; we then scan the Wi-Fi frequency for P_{on} iterations, and build the corresponding set $|AP_u^{on}|$. Similar to (1), we store a record $R_j^{on} = \langle BSSID_j^{on}, RSS_j^{on}, \mu_j^{on}, \sigma_j^{on} \rangle, \forall j \in AP_u^{on}$. The localization process is performed by a pattern matching algorithm, which compares the records R_{ij}^{off} and R_j^{on} , and determines the RP (denoted as RP^*) corresponding to the expected user location. In this study, we consider the following pattern-matching algorithms, which have been also implemented in Wi-LO (Section IV) and evaluated on test-beds (Section V):

- **BSSID Based (BB)**. The algorithm computes a score function $S(i)$ for each RP i defined as the number of APs that have been detected in both AP_u^{on} and AP_i^{off} , i.e. $S(i) = |AP_i^{off} \cap AP_u^{on}|$. The RP with the highest score is returned as the expected location, i.e. $RP^* = \operatorname{argmax} S(i), \forall i \in P_{rp}$. If there are multiple RPs with equal highest score, a random one is picked and returned.
- **Nearest Neighbour (NN)**. The algorithm computes a score function $S(i)$ for each RP i , defined as the Euclidean distance between average RSS values at RP i and at u , for all the APs that have been detected in both AP_u^{on} and AP_i^{off} , i.e.:

$$S(i) = \sqrt{\sum_j (\mu_{ij}^{off} - \mu_j^{on})^2}, \forall j \in AP_i^{off} \cap AP_u^{on} \quad (2)$$

Moreover, if $|AP_i^{off}| > |AP_u^{on}|$, we add to each $S(i)$ an extra penalty accounting for the BSSID-based error, i.e.

$S(i) = S(i) + (|AP_i^{off}| - |AP_u^{on}|) \cdot \gamma$, with γ being a constant value. The RP with the lowest error is returned as the expected location, i.e. $RP^* = \operatorname{argmin} S(i), \forall i \in P_{rp}$.

- **Likelihood Estimator (LE)**. The algorithm evaluates the Gaussian Likelihood $L(i, j)$ of the RSS samples for each AP j that have been detected in both AP_u^{on} and AP_i^{off} . The value of $L(i, j)$ is computed as follows:

$$L(i, j) = \Pi_{l=0}^{|RSS_u^{on}|} \frac{1}{\sqrt{2\pi\sigma_{ij}^{off}}} \cdot \exp \left\{ \frac{(RSS_u^{on}[l] - \mu_{ij}^{off})^2}{2 \cdot \sigma_{ij}^{off}} \right\} \quad (3)$$

Then, the algorithm computes a score function $S(i)$ for each RP i , summing up all the $L(i, j)$ values for the APs j detected in both AP_u^{on} and AP_i^{off} , i.e. $S(i) = \sum L(i, j) \forall j \in AP_i^{off} \cap AP_u^{on}$. An extra penalty accounting for the BSSID-based error is introduced, as for the NN algorithm. The RP with the highest score is returned as the expected location, i.e. $RP^* = \operatorname{argmax} S(i), \forall i \in P_{rp}$.

- **Welch's t-test (WT)**. The algorithm computes the Welch's t-test for each AP j that is detected in both AP_u^{on} and AP_i^{off} . The WT allows taking into account the similarity between the mean values, i.e. μ_{ij}^{off} and μ_j^{on} , when computed over populations of different sizes, i.e. $P_{on} \neq P_{off}$. We evaluate the null hypothesis $H_0(j)$ where average RSS values are equal for AP j , i.e. $\mu_{ij}^{off} \approx \mu_j^{on}, \forall j \in AP_i^{off} \cap AP_u^{on}$ and $i \in P_{rp}$. Then, the algorithm computes a score function $S(i)$ for each RP i , defined as the number of APs satisfying the null hypothesis, i.e. $S(i) = |\{j | H_0(j) = \text{true} \forall j \in AP_i^{off} \cap AP_u^{on}\}|$. Finally, the RP with the highest score is returned as the expected location, i.e. $RP^* = \operatorname{argmax} S(i), \forall i \in P_{rp}$.

For space reasons, we report only a subset of the simulation results with the following setting of the parameters: $P_{nside}=150$ meters, $P_{tside}=50$ meters, $P_{rside}=5$ meters, $P_{rpside}=5$ meters, $P_{nfloors}=5$, $P_\sigma=6\text{db}$. We consider an attenuation factor of 12dbm for each traversed outside wall, and 5dbm per each indoor wall. Each data point is computed by averaging the results of 10000 runs. Within each run, we randomly generate the user position, the locations of the P_{nap} APs within the scenario, and the shadowing variance ($P_{i,\sigma} < P_\sigma$) at each RP i . We restrict the analysis to a 2D plane, i.e. we assume the knowledge of the floor where the user is currently located. This is in line with recent studies that show the elevation from floor-level can be efficiently detected by the barometer sensor without Wi-Fi radio fingerprinting [23]. Figure 2(a) depicts the accuracy of RP detection, by varying the density of APs in the scenario, for the four pattern matching algorithms previously introduced. We can see that: (i) for $P_{nap} < 10$, the RP detection accuracy is lower than 40% for all the algorithms; (ii) the WT approach slightly outperforms the NN and LE; (iii) the accuracy increases by adding more RPs till a maximum value, after which adding more APs does not produce significant gains. This aspects is also confirmed by the heat-map of Figure 2(b), showing the average localization

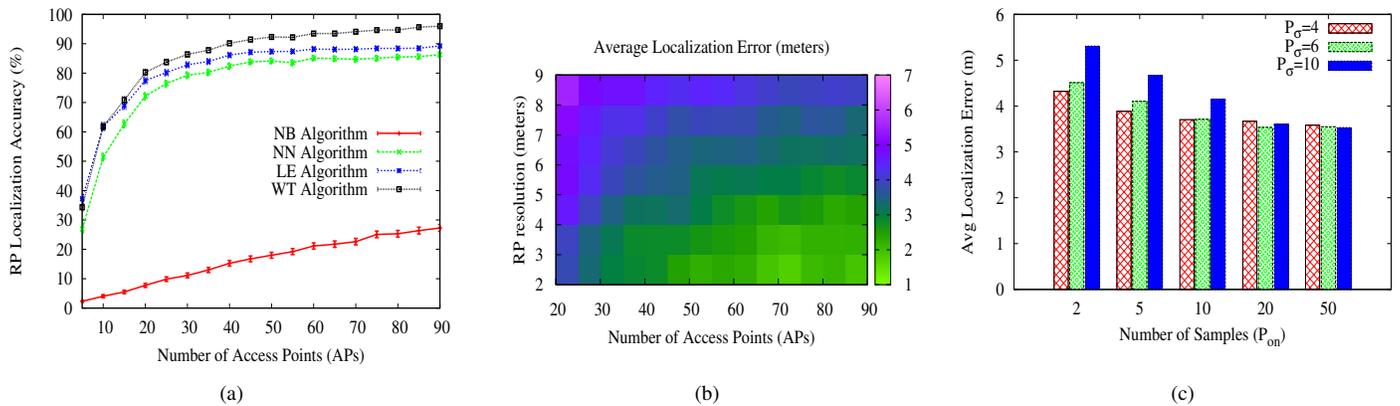


Fig. 2. Impact of radio fingerprinting parameters on the localization performance: AP density (Figure 2(a)), RP granularity (Figure 2(b)) and number of detection samples (Figure 2(c)).

error (color of the cell) as a function of P_{nap} (x -axis) and the RP granularity P_{rpside} (y -axis), for the WT algorithm. Reducing the P_{rpside} reduces the average error, at the cost of increasing the complexity of the training phase. At the same time, fixing a specific P_{rpside} value, there exists an optimal minimum P_{nap}^* guaranteeing also the minimum localization error. Finally, Figure 2(c) investigates the relationship between the length of the duration phase (P_{on} on the x -axis) and the localization error of the WT algorithm, for different values of P_{σ} . Increasing the detection time (P_{on}) can reduce the localization error, although adding to the localization lag. Moreover, it might not always be possible in presence of user's mobility. The optimal value of P_{on} is clearly affected by the propagation conditions, i.e. by P_{σ} . At the same time, given P_{σ} , there is an optimal minimum detection time P_{on}^* , such that increasing it does not improve the gain achieved in terms of localization precision.

IV. SOFTWARE ARCHITECTURE

Based on our simulation study, we describe data-fusion techniques that aim at coping with the variable density of Wi-Fi APs at a given location. To this aim, we built a software platform for indoor localization, called Wireless Locator (Wi-LO), whose client-server architecture is depicted in Figure 3.

Wi-LO supports four different input sources of radio fingerprinting: Wi-Fi, LTE, BLE, MAG, while the barometer might be optionally employed to detect the user floor. The choice of which input source to consider for the map building during the offline phase, and which for localization during the online phase, is left to the user through the GUI of the mobile app. Let I_{off} and I_{on} be the sets of selected input sources, with $I_{on} \subseteq I_{off}$. More in details, the client software (implemented for Android devices) is in charge of performing measurements at specific RPs, and transmitting them to the server; here, the RF maps are aggregated and stored in a NO-SQL Mongo-DB database. Moreover, the server side is in charge of performing the localization process, by running the selected pattern matching algorithms on each selected source in I_{on} , then running the fusion scheme, and finally returning

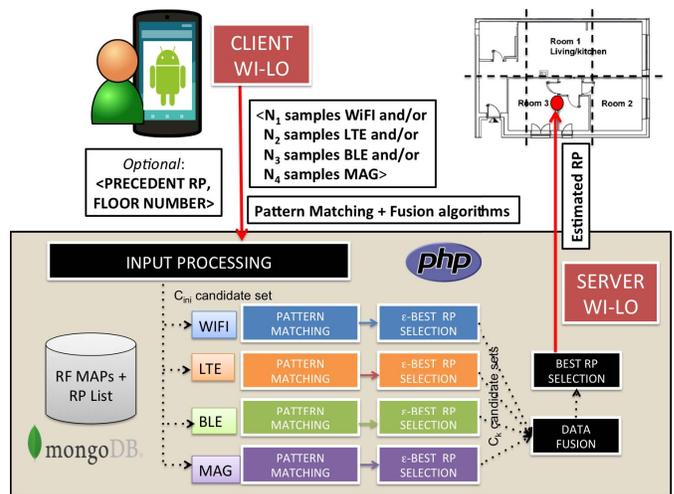


Fig. 3. The Wi-LO software architecture.

the expected RP^* to the client. Again, pattern matching and fusion algorithms are decided by the user through the Android application running on the smart device. All the localization algorithms have been implemented in PHP, by making large utilization of class inheritance and interfaces, in order to maximize the modularity and extensibility of the tool. During the online phase, the client performs the following actions:

- 1) It gathers Wi-Fi and/or LTE and/or BLE and/or MAG samples for T_d seconds, based on the sources selected in I_{on} . For each detected Wi-Fi AP and BLE beacon, it builds the record from (1). For BLE and MAG measurements, a single record is built without the BSSID field. For the tri-axial MAG, a synthesized magnitude value [24] is extracted, at a fixed frequency (1 value/sec).
- 2) Optionally, it also detects the current floor by sampling the barometer sensor, and applying the technique described in [23].
- 3) It then transmits the gathered data records to the server, together with the floor location and the previous esti-

mated RP (in case of trajectory tracking).

The server software processes the input and computes the expected RP^* as follows:

- 1) First, it builds an initial set of candidate RPs C_{ini} , based on the floor location and the previous estimated RP, if available. Otherwise, $C_{ini} = P_{rp}$.
- 2) Then, it applies the selected pattern matching algorithm on each source. At present, we implemented the four algorithms described in Section III. Each algorithm assigns a score $S_k(i)$ to each RP $i \in C$ and data source k . Next, the ϵ -Best RP Selection module is executed, which computes the maximum score $Max(S_k)$, and populates the candidate set C_k for source $k \in I_{on}$ with all the RPs including a relative score difference with $Max(S_k)$ lower than ϵ , i.e. $C_k = \{RP\ j\ s.t.\ |\frac{S_k(i) - Max(S_k)}{Max(S_k)}| \leq \epsilon\}$. We also define the confidence metric of source k (denoted as $\kappa(k)$) as the inverse of the cardinality of the candidate set, i.e. $\kappa(k) = \frac{1}{|C_k|}$.
- 3) If a single source k has been selected, than no data fusion is applied, and the RP with score equal to $Max(S_k)$ is returned. Otherwise, we merge the results of the candidate sets C_k . Let $U_k(i)$ be the identity function returning 1 if RP $i \in C_k$, 0 otherwise. The Fusion Module of Figure 3 works by computing a Fusion Score to each RP $i \in P_{rp}$, based on this Equation:

$$FS(i) = \sum_{k=1}^{|I_{on}|} U_k(i) \cdot \omega_k(i) \quad (4)$$

where $0 \leq \omega_k(i) \leq 1$ is the weight applied to the decision of source k , and depends on the specific fusion scheme selected for localization.

- 4) Finally, the Best Selection Module returns RP^* maximizing the FS function. A random selection is applied in case of multiple choices for the maximum value.

In this paper, we focus on a specific novel fusion-scheme, called *diversity-based* scheme, which is evaluated and compared against two competing approaches, namely the majority and weight-based algorithms. We briefly illustrate the three fusion schemes in the following:

Diversity-based Approach. The rationale behind this algorithm is that different sources might experience different spatial resolution during the offline/online phases. For instance, it is well known that LTE RSS samples at different indoor locations can exhibit smaller differences than Wi-Fi RSS samples: moreover, on several smart devices, the granularity in the LTE RSS can be limited to a small range of values [15]. At the same time, there is the chance to observe significant signal diversities which can be exploited as location-specific signatures. Based on such intuition, we compute a Diversity Metric value (DM) between each couple of RPs of the scenario, for each source. The $DM_k(i, j)$ value between RP i and j is defined as follows:

$$DM_k(i, j) = \frac{|S_k(i) - S_k(j)|}{|max(S_k(i), S_k(j))|} \quad (5)$$

where $S_k(i)$ is the score function applied by the selected pattern matching algorithm (i.e. BB, NN, LE, WT) on source k and for RP i . Details on how the $S_k(\cdot)$ values are computed for each specific pattern matching algorithm have been provided in Section III. We just remark here that the DM metric: (i) is agnostic of the specific pattern matching algorithm in use, (ii) assumes values in range [0,1]. We then define the Diversity Metric of RP i on source k , i.e. $DM_k(i)$, as the average $DM(i, j)$ considering all the other RPs j of the scenario, i.e.:

$$DM_k(i) = \frac{1}{|P_{rp}|} \sum_{j=1, i \neq j}^{P_{nrp}} DM_k(i, j) \quad (6)$$

Then, we compare the DM values of different sources, and we assign the $\omega_k(i)$ values proportionally to $DM_k(i)$, i.e.:

$$\omega_k(i) = \frac{DM_k(i)}{\sum_{l=1}^{|I_{on}|} DM_l(i)} \quad (7)$$

The $\omega_k(i)$ values are pre-computed at the end of the training phase, and stored within the database on the server, together with the RF maps. In Section V, we provide further insights about the behaviour of the Diversity-based scheme, by plotting the $\omega_k(i)$ values at different locations of a test-bed scenario.

Majority-based. This scheme does not introduce any difference between the sources, i.e. $\omega_k(i) = 1 \forall k \in I_{on}$. As a result, the RP^* indicated by more sources as the expected location is returned to the user.

Confidence-based. This scheme assigns weights to each source based on the confidence computed by the ϵ -Best Selection Module, i.e. $\omega_k(i) = \kappa(k) \forall k \in I_{on}$. The rationale here is that the sources exhibiting less uncertainty about the user's localization should gain more trust.

V. EXPERIMENTAL RESULTS

In this section, we provide experimental results on the indoor localization capabilities provided by Wi-LO. We consider two buildings for testing, located at two different geographical areas, and characterized by different AP densities: (i) an urban-scenario, where the number of APs detected at each RP is always between the 15 and 25 units; and a (ii) a rural-scenario, where the number of APs detected at each RP is always below the 3 units. In both the scenarios, the indoor area is approximatively equal to 225 m². Although supported by Wi-LO, we did not consider BLE among the localization sources used for our tests. This is because we were interested in understanding the possibility to localize an user relying on the infrastructures available on site (e.g. external APs and cellular network), without any cost for the set-up; clearly, the deployments of BLE beacons within the target scenario might increase the system performance. For each scenario and algorithm, we executed over 30 tests, and then computed the average accuracy in detecting the RP where the user is located. Figures 4(a), 4(b) and 4(c) refer to the configuration with $P_{rpside}=5$ meters. More specifically, Figure 4(a) and 4(b) show the average accuracy of different sources and pattern matching algorithms, for the urban (Figure 4(a)) and rural

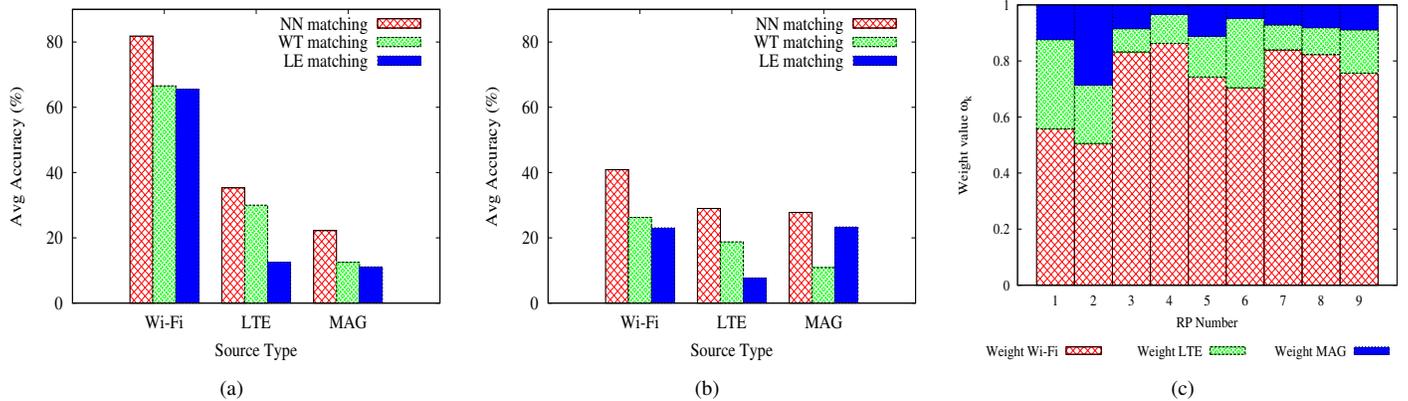


Fig. 4. The RP detection accuracy for the urban and rural scenarios are shown in Figures 4(a) and 4(b), respectively. The weight distribution at different RPs for the Diversity-based fusion scheme is depicted in Figure 4(c).

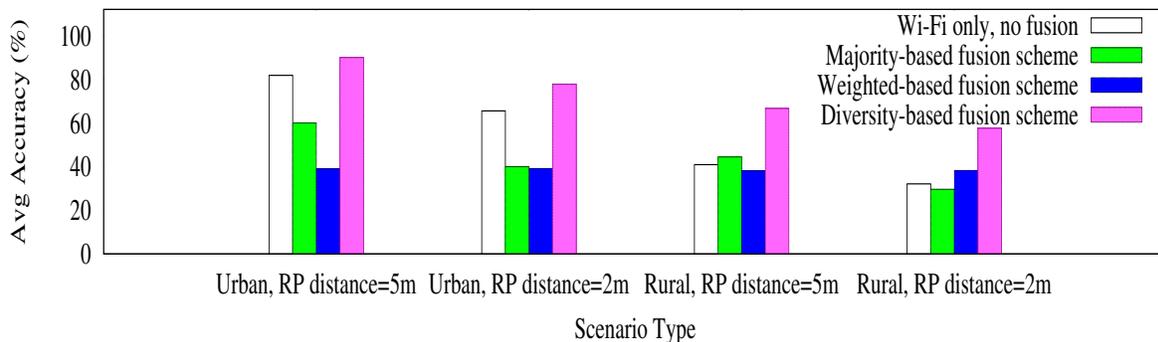


Fig. 5. The RP detection accuracy for different scenarios, RP granularity and fusion schemes.

scenarios (Figure 4(b)). Both the figures confirm that Wi-Fi radio fingerprinting is -on average- more accurate than the LTE and MAG equivalents, although its performance strongly decreases from the urban to the rural scenario (while the LTE and MAG are slightly affected). At the same time, the figure shows that the NN pattern matching scheme outperforms the LE and WT, for all sources and in both scenarios. This can be explained in several ways: (i) for LTE and MAG sources, the RSS variance at a specific RP might not be significant, since the signal might present no fluctuations over time (LTE), or fluctuations might be limited in magnitude (MAG); (ii) for the Wi-Fi sources, the assumption of normality of signal distribution may not completely hold at each RP, or a long number of samples should be gathered during the online phase. Next, we investigate the possible gain achieved by data fusion approaches combining hard decisions of Wi-Fi, MAG and LTE radio fingerprinting schemes. Figure 5 shows the average localization accuracy for the three fusion schemes described in Section IV for the urban and rural scenarios, where we also vary the RP granularity ($P_{rpside}=5$ meters and $P_{rpside}=2$ meters). All the fusion schemes employ the NN as pattern matching scheme on each source, since it has been shown to provide the best performance in both scenarios. In Figure 5, we also report the accuracy of the Wi-Fi radio fingerprinting

scheme without data fusion (basically the Wi-Fi NN bars of Figures 4(a) and 4(b)). It is remarkable to notice that the Majority and Granularity-based fusion scheme perform similarly, and in most cases worse than the pure Wi-Fi scheme; this can be justified since we have previously shown that -on average- the MAG and LTE sources are less accurate than the Wi-Fi. We can conclude that merging the hard decisions of different sources does not lead automatically to a performance increase. At the same time, Figure 5 shows that the Diversity-based approach overcomes both the pure Wi-Fi and the other fusion schemes, in all the configurations tested. This is due to Equation 7, which accounts for the utilization of Wi-Fi, MAG and LTE sources based on their effectively ability to recognize a radio fingerprint at a specific location. To this aim, Figure 4(c) depicts the values of ω_k for the Urban scenario and $P_{rpside}=5$, at different RPs (on the x -axis). We can see that the weights ω_k change from location to location, although the Wi-Fi source always gets the highest values in accordance with Figure 4(a). Averaging over the P_{rpside} values, the Diversity-based scheme achieves a performance gain of +8% on urban scenarios, and +25% on rural scenarios, compared to the pure Wi-Fi scheme. Finally, we conclude by reporting the average localization lag over the Rural scenario, for different algorithms. The localization lag is defined as

TABLE I
LOCALIZATION LAG

Algorithm	Average time (seconds)
NN, Wi-Fi only	0.439
WT, Wi-Fi only	0.416
LE, Wi-Fi only	0.515
Diversity-based	0.591

the time required for computing and returning the expected localization output (i.e. RP^*) to the client device; it includes the overhead involved by the client-server communication and by the data processing and algorithm execution at the server side, but it does not include the detection sampling time (T_{on}), since this parameter is decided by the user. We see that the Diversity-based scheme does not introduce much significant overhead to the localization process, since most of the time delay depends from the client-server communication.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we have addressed the problem of indoor localization of smart devices through the design, implementation and evaluation of a software paradigm called Wi-LO, which maximally exploits all the available radio interfaces. Wi-LO fuses information from multiple sources (i.e. Wi-Fi, LTE, BLE and MAG), supports several different pattern matching algorithms (BB, NN, LE and WT) on each source, and hard fusion schemes combining the decisions of each source. We have experimentally evaluated the performance of fusion and no-fusion algorithms for two scenarios, characterized by different density of external APs. Results show that the fusion schemes based on simple voting schemes might not improve the localization performance, because of the different resolution of the Wi-Fi, LTE and MAG radio maps. At the same time, a diversity-based scheme taking into account the reliability of each source can effectively improve the performance of the localization process, compared to a pure Wi-Fi scheme. Future works include: the integration within Wi-LO of dead-reckoning techniques based on smart device's embedded sensors (i.e. accelerometer/gyroscope), the design of additional mechanisms for the automatic selection of pattern matching and fusion algorithms, the testing in multi-floor scenarios.

VII. ACKNOWLEDGMENT

This work was supported in part by the U.S. Office of Naval Research under grant number N00014-16-1-2651.

REFERENCES

[1] H. Liu and H. Darabi. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on systems, man and cybernetics*, pp. 1067-1080, 2007.

[2] K. Subbu, C. Zhang, J. Luo and A. V. Vasilakos. Analysis and status quo of smartphone-based indoor localization systems. *IEEE Wireless Communications*, pp. 106-112, 2014.

[3] B. Wang, Q. Chen, L. T. Yang and H.-C. Chao. Indoor smartphone localization via fingerprint crowdsourcing: challenges and approaches. *IEEE Wireless Communications*, pp. 82-89, 2016.

[4] P. Bahl and V.N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. *IEEE INFOCOM*, Tel Aviv, Israel, 2000.

[5] S. Yang, P. Dessai, M. Verma, and M. Gerla. FreeLoc: Calibration-free crowdsourced indoor localization. *IEEE INFOCOM*, Turin, Italy, 2013.

[6] L. Chen, K. Yang and X. Wang. Robust cooperative Wi-Fi fingerprinting-based indoor localization. *IEEE Internet of Things*, pp. 1406-1417, 2016.

[7] Y. Kim, Y. Chon and H. Cha. Mobile Crowdsensing Framework for a Large-Scale Wi-Fi Fingerprinting System. *IEEE Pervasive Computing*, pp. 58-67, 2016.

[8] C. Wu, Z. Yang, C. Xiao, C. Yang, Y. Liu and M. Liu. Static power of mobile devices: self-updating radio maps for wireless indoor localization. *Proc. of IEEE INFOCOM*, Hong-Kong, China, 2015.

[9] D. Mascharka and E. Manley. LIPS: Learning based indoor positioning system using mobile phone-based sensors. *Proc. of IEEE CCNC*, Las Vegas, USA, 2016.

[10] E. S. Lahan, J. Talvitie and G.-S. Granados. Data Fusion Approaches for WiFi Fingerprinting. *Proc. of IEEE ICL-GNSS*, Barcelona, Spain, 2016.

[11] M. Bosling, M. Ceriotti and T. Zimmermann. Fingerprinting channel dynamics in indoor low-power wireless networks. *Proc. of IEEE Wintech*, Miami, USA, 2013.

[12] X. Tian, R. Shen, D. Liu, Y. Wen and X. Wang. Performance analysis of RSS fingerprinting based indoor localization. *IEEE Transactions on Mobile Computing*, pp. 1-14, 2016.

[13] T. Guo, Y. Li and Y. Sun. Accurate indoor localization based on crowd sensing. *Proc. of IEEE IWCWC*, Paphos, Cyprus, 2016.

[14] R. Zhang, A. Bannoura and F. Hoffinger. Indoor localization using a smart phone. *Proc. of IEEE SAS*, Galveston, USA, 2013.

[15] E. Martin. Multimode radio fingerprinting for localization. *Proc. of IEEE RWS*, Santa Clara, USA, 2011.

[16] J. Turkka, T. Hiltunen, R. U. Mondal and T. Ristaniemi. Performance evaluation of LTE radio fingerprinting using field measurements. *Proc. of IEEE ICICS*, Singapore, 2015.

[17] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li and F. Zhao. Magicol: indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE Journal on Selected Areas on Communications*, pp. 1443-1457, 2015.

[18] F. Gu, J. Niu, and L. Duan. WAIPO: A fusion-based collaborative indoor localization system on smartphones. *IEEE/ACM Transactions on Networking*, pp. 1-14, 2017.

[19] T. Banerjee, K. R Chowdhury, and DP Agrawal. Using polynomial regression for data representation in wireless sensor networks. *International Journal of Communication Systems*, 20 (7), 829-856, 2007.

[20] B. Kim, M. Kwak and J. Lee T. T. Kwon. A multi-pronged approach for indoor positioning with Wi-Fi, magnetic and cellular signals. *Proc. of IEEE IPIN*, Busan, Korea, 2014.

[21] P. Mirowski, T. K. Ho, S. Yi and M. MacDonald. SignalSLAM: Simultaneous localization and mapping with mixed Wi-Fi, Bluetooth, LTE and magnetic signals. *Proc. of IEEE IPIN*, Montbeliard-Belfort, France, 2013.

[22] M. Di Felice, R. Doost-Mohammady, K. R. Chowdhury and L. Bononi. Smart Radios for Smart Vehicles: Cognitive Vehicular Networks. *IEEE Vehicular Technology Magazine*, pp. 26-33, 2012.

[23] L. Bedogni, F. Franzoso, L. Bononi. A Self-Adapting Algorithm Based on Atmospheric Pressure to Localize Indoor Devices. *Proc. of IEEE GLOBECOM*, Washington, USA, 2016.

[24] L. Bedogni, M. Di Felice, L. Bononi. By Train or By Car? Detecting the User's Motion Type through Smartphone Sensors Data. *Proc. of IEEE Wireless Days*, Dublin, Ireland, 2012.