

Toward Extrapolation of WiFi Fingerprinting Performance Across Environments

Filip Lemic
Technische Universität Berlin
lemic@tkn.tu-berlin.de

Vlado Handziski
Technische Universität Berlin
handziski@tkn.tu-berlin.de

Giuseppe Caso
Sapienza University of Rome
caso@newyork.ing.uniroma1.it

Pieter Crombez
Televic Health Care NV
p.crombez@televic.com

Luca De Nardis
Sapienza University of Rome
lucadn@newyork.ing.uniroma1.it

Adam Wolisz
Technische Universität Berlin
awo@ieee.org

ABSTRACT

Out of the plethora of approaches for indoor localization, WiFi-based fingerprinting offers attractive trade-off between deployment overheads and accuracy. This has motivated intense research interest resulting in many proposed algorithms which are typically evaluated only in a single or small number of discrete environments. When the end-user's environment is not part of the evaluated set, it remains unclear if and to what extent the reported performance results can be extrapolated to this new environment. In this paper, we aim at establishing a relationship between the similarities among a set of different deployment environments and parameterizations of fingerprinting algorithms on one side, and the performance of these algorithms on the other. We hypothesize about the factors that can be used to capture the degree of similarity among environments and parameterizations of the algorithms, and proceed to systematically analyze the performance of two fingerprinting algorithms across four environments with different levels of similarity. The results show that the localization error distributions have small statistical difference across environments and parameterizations that are considered similar according to our hypothesis. As the level of similarity is decreased, we demonstrate that the relative performance of the algorithms can still be preserved across environments. For dissimilar environments, the localization errors demonstrate larger statistical differences.

Keywords

Indoor localization; indoor positioning; fingerprinting; performance extrapolation; radio frequency; WiFi.

1. INTRODUCTION

Radio Frequency (RF)-based localization, specifically WiFi Received Signal Strength Indicator (RSSI)-based fingerprinting, is one of the promising candidates for an ubiquitous indoor localization service [1]. Consequently, performance evaluations of such algorithms in different environments are becoming publicly available [2, 3]. Thus, the interested users

are able to get insights in the performance of various algorithms in different environments through these standardized benchmarks. However, it is still unclear if the performance results, in terms of localization errors, achieved in one environment and for one parameterization of an algorithm, can be representative for another environment and parameterization of the same algorithm.

The extrapolation of the performance of fingerprinting algorithms across environments would be beneficial from two perspectives. From the technological and research perspective, it would allow categorization of algorithms based on their suitability for different types of environments. From the users' perspective, the extrapolation would give them the possibility of speculating about the performance of different algorithms for a particular environment without the need for extensive experimentation. Collection of such measurements is particularly problematic for buildings that are usually not accessible for an extensive experimentation (e.g. hospitals or buildings under construction).

The possibility of performance extrapolation for different systems has been addressed in various research domains. For example, in network experimentation the need for having realistic experimentation conditions in testbeds has been addressed in [4], while for evaluation of artificial intelligence the question of extrapolation of results achieved in testbeds to reality has been addressed in [5]. Moreover, in [6] the authors aim on predicting the performance of GPU applications by correlating them to existing benchmarks, while in [7] the authors qualify the similarity of computer systems and then use this similarity for predicting the performance of a new application. However, in the domain of WiFi fingerprinting, the question of performance extrapolation is, to the best of our knowledge, still open.

In this paper, we systematically analyze the performance achieved by a set of WiFi RSSI-based fingerprinting algorithms in a set of environments with different characteristics. Our goal is to establish a link between the similarities among environments and parameterizations of algorithms in these environments, with the possibility to extrapolate the performance of such algorithms across environments. Our contributions include a hypothesis about the extrapolability of the performance of WiFi fingerprinting algorithms across environments, demonstrating the feasibility of the hypothesis, and outlining a methodology for its further evaluation.

This paper is structured as follows. In Section 2, we formulate our hypothesis about the possibility of performance extrapolation of fingerprinting algorithms across environments. We characterize the similarity among environments

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HotMobile'16, February 23–24, 2016, St. Augustine, FL, USA

© 2016 ACM. ISBN 978-1-4503-4145-5/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2873587.2873588>

by looking at two aspects: physical shape and setup, and RF propagation characteristics. In Section 3, we describe the physical shape and experimental setup in all environments used in the evaluation. In Section 4, we introduce a model used for calculating propagation characteristics of different environments. In Section 5, we overview the evaluated algorithms and provide the evaluation scenarios and approaches, while in Section 6 we discuss the obtained results. Finally, in Section 7, we outline directions for further evaluation of the given hypothesis and we conclude the work.

2. EXTRAPOLATION HYPOTHESIS

For the accurate extrapolation of the performance of fingerprinting algorithms across environments one has to be careful to select both environments and parameterizations of an algorithm that can be characterized as “similar”. One of the important parameters in accessing the similarity of two environments pertains to their physical shape, *w.r.t.* their outer size, as well as the sizes of their inner spaces. We further hypothesize that similarities in the parameters of propagation in different environments are important factors for characterizing the similarity among environments.

Apart from the environments *per-se*, we hypothesize that the parameterizations of a fingerprinting algorithm in these environments have to be similar for being able to extrapolate the performance of an algorithm across environments. This pertains to the type of algorithm used, the number, density and deployment locations of Access Points (APs), their transmission powers and operating frequencies, the number of measurements taken at each measurement point, and the number and density of training points. Given that the parameterizations of an algorithm in two environments are comparable and the environments are similar, we hypothesize that a reliable extrapolation of the algorithm’s performance across environments is possible.

We evaluate the hypothesis by comparing the performance achieved by two fingerprinting algorithms in a baseline, real-life hospital environment with their performance in three other environments with different levels of similarities to the hospital environment. The performance results of fingerprinting algorithms in these additional environments are publicly available [8], and they serve as standardized benchmarks of the performance of fingerprinting algorithms. Our similarity characterization yields that one environment is highly similar to the hospital baseline, in both physical and propagation characteristics, while the other two environments have increasing dissimilarities with the baseline.

3. ENVIRONMENTS: PHYSICAL SHAPE AND EXPERIMENTAL SETUP

In this section, we present the physical shape and the experimental setup in the four used environments, with the relevant environment related parameters given in Table 1.

Hospital: The hospital serves as a baseline environment for which we would like to extrapolate the performance of fingerprinting algorithms. The measurement campaign was performed in the “chirurgic day” ward of the Sint-Jozefskliniek Izegem in Belgium, with footprint depicted in Figure 1a). For the measurement collection the end section of a corridor was used, while the rest of ward was in “normal operation”, meaning people were present in the hospital. In the given environment, six IEEE 802.11g APs (WNDR 4300) were deployed, one in each room in the environment of interest, with

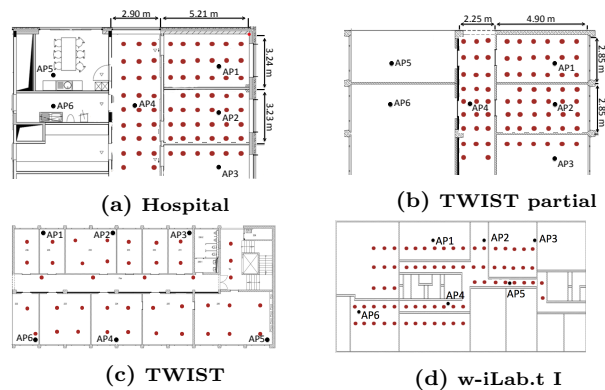


Figure 1: Footprints of the four environments

their locations indicated in the figure. The APs were configured in beaconing mode, with beacon transmission period of 100 ms. The transmission power of each AP was set to 20 dBm. The APs 1, 3 and 4 were operating on WiFi channel 6, while for the others (AP 2, 5 and 6) the operational channel was WiFi channel 11. In the environment of interest, which included three rooms and a hallway, a total number of 73 measurement points were defined in a relatively dense grid with cell size equal to 1 m, as indicated by the red dots in Figure 1a). In each of the defined points, the wireless environment was sampled for WiFi beacon packets using a laptop with external wireless adapter (TL-WN823N). At each location, four scans of the wireless environment were performed, meaning that at each location, from each visible AP, at most 4 beacon packets could be obtained. Due to a limited availability of the environment and in order not to interrupt the normal hospital activities, the measurement campaign was constrained in both size and duration. However, it represents one of the first publicly available datasets that capture a realistic hospital environment for the purpose of evaluation of WiFi-based fingerprinting.

TWIST partial: The standardized benchmark obtained from this environment, which is a part of the TWIST testbed in Berlin [9], has high levels of similarity with the hospital environment, *w.r.t.* similar sizes of the offices/rooms, outer sizes, and similar activities as in the hospital (people moving in the testbed premises, since it is an office building). We believe that this environment mimics the hospital environment to the level that is practically possible, as shown in Figure 1b). In this part of the environment, six WiFi APs were deployed at similar locations as for the hospital. The APs were configured as in the hospital, although they were of a different type (TP LINK N750). Similarly, 73 measurement points were selected and in each point the wireless environment was scanned four times, same as for the hospital, but using an Airport Extreme network interface card.

TWIST: The TWIST testbed environment in its entirety is an office environment, with room sizes slightly bigger than the hospital environment and with an outer dimensions of roughly $30 \times 15 m^2$. Also in this environment, six dedicated WiFi APs were deployed with their locations shown in Figure 1c). They were configured to operate in beaconing mode with beacon transmission period of 100 ms and transmission power of 20 dBm. In this environment, 41 measurement points were defined and in each of the points four scans of the wireless environment were performed.

w-iLab.t I: The w-iLab.t I testbed is an office environment with outer dimensions of roughly $45 \times 17 m^2$. In comparison to the TWIST testbed, in this environment the office

sizes are less similar to the hospital, as shown in Figure 1d). For this measurement campaign also six “dedicated” WiFi APs were deployed with the same configuration as in the TWIST environment. For the measurement campaign 69 measurement points were defined and in each of them a scan of the wireless environment was performed. In this environment also four measurements per point were taken.

Table 1: Parameters of the environments

Parameter	Outer size	Mean room size	Wall type
Hospital	8.1 x 7.4 m	15.0 m ²	plywood
TWIST partial	7.2 x 6.8 m	12.5 m ²	concrete
TWIST	30.0 x 15.0 m	27.0 m ²	concrete
w-iLab.t I	45.0 x 17.0 m	51.0 m ²	plywood

4. PROPAGATION CHARACTERISTICS

For further understanding of the similarities among environments, we modeled the RF propagation parameters by leveraging the collected measurements. The applied propagation model is the COST 231 multi-wall model for indoor radio propagation [10]. The applicability of this model for indoor localization purposes has been demonstrated in [11]. The first attenuation contribution in the model is a well-known one-slope term that relates the received power to the distance. Two parameters influence the attenuation in this term: the constant l_0 (the path-loss at 1 m distance and at the center frequency of 2.45 GHz) and the path-loss exponent γ . The second attenuation contribution is a linear wall/obstacle term. The number of obstacles in the direct path between transmitter and receiver is counted and for each type of obstacle an attenuation contribution is assumed. Given the model and the site-specific measurements from different environments, we leveraged a least square fitting procedure that allows minimizing the cost function between the measured received power and the modeled one.

Table 2: Parameters of the propagation model

Parameter	l_c [dB]	γ	l_w [dB]
Hospital	57.38	1.46	2.60
TWIST partial	58.36	1.25	2.78
TWIST	53.73	1.64	4.51
w-iLab.t I	60.23	1.29	1.12

The parameters to be optimized are the constant l_c related to the least square fitting procedure, the γ path-loss exponent, and the wall attenuation factor. Although $\gamma = 2$ is a usual assumption for the propagation in the free space, coming from Friis equation, due to the obstacles in our environments we also estimate γ . Moreover, we do a set of prediction tests on (rolling) 10% of sampling points using a model fitted on the remaining 90%, to be sure that we do not have unexpected interactions between the linear constant and the γ . The modeled averaged propagation parameters are given in Table 2, since small discrepancies are obtained by performing the rolling tests. As indicated in the table, for all environments l_c and γ variables have comparable values. Furthermore, in terms of wall attenuation, the similarity between the hospital and the partial TWIST environments is high, while smaller similarities are observed between the hospital, TWIST and w-iLab.t I environments.

5. EVALUATED ALGORITHMS, SCENARIOS AND APPROACHES

This section gives an overview of the fingerprinting algorithms, scenarios, and approaches used in the evaluation.

5.1 Fingerprinting Algorithms

Euclidean distance of averaged RSSI vectors: This simple, yet popular fingerprinting algorithm [12] computes an average value of the RSSI measurements obtained from each AP used for localization. The fingerprint is a vector of average values of the RSSI measurements obtained from all APs used for localization in both training and runtime steps, where K is the length of the vector. Let $\bar{\mathbf{X}}_{t,m} = [\overline{RSSI}_{t,1}, \dots, \overline{RSSI}_{t,k}, \dots, \overline{RSSI}_{t,K}]$ be the vector of averaged RSSI values $\overline{RSSI}_{t,i}$ from each AP i obtained in training step at point $m \in 1, \dots, M_t$, i.e. training fingerprint. In the same manner, let $\bar{\mathbf{X}}_r = [\overline{RSSI}_{r,1}, \dots, \overline{RSSI}_{r,k}, \dots, \overline{RSSI}_{r,K}]$ be the vector of averaged RSSI values $\overline{RSSI}_{r,i}$ from each AP i obtained in runtime step, i.e. runtime fingerprint. The pattern matching procedure uses Euclidean distance between a training fingerprint at the cell m and the runtime fingerprint:

$$D_E(\bar{\mathbf{X}}_{t,m}, \bar{\mathbf{X}}_r) = |\bar{\mathbf{X}}_{t,m} - \bar{\mathbf{X}}_r|. \quad (1)$$

Training fingerprints with the smallest distance D_E are then used in a post-processing procedure.

Pompeiu-Hausdorff distance of RSSI quantiles: A recently proposed procedure [12] uses a vector of q quantiles of the RSSI values from each AP as fingerprints, which are calculated in two steps. First the Cumulative Distribution Function (CDF) of the RSSI measurements from each AP is computed. Second, the quantiles, i.e. RSSI values with probabilities $k/(q-1)$, where $k = 0, 1, \dots, q-1$, are calculated. The result of the quantile calculation in both training and runtime steps is a quantile matrix $Q_{K,q}$, where K is the number of APs visible at the given location and q is a number of quantiles. The pattern matching procedure of this algorithm uses the Pompeiu-Hausdorff (PH) metric for capturing similarities between training fingerprints and a runtime fingerprint, with $d(x_{t,k}, x_{r,k})$ being the Euclidean distance between elements of the runtime fingerprint \mathbf{X}_r and training fingerprint $\mathbf{X}_{t,m}$ at point m :

$$D_{PH}(\mathbf{X}_{t,m}, \mathbf{X}_r) = \max_{x_{t,k} \in \mathbf{X}_{t,m}} \min_{x_{r,k} \in \mathbf{X}_r} d(x_{t,k}, x_{r,k}). \quad (2)$$

5.2 Evaluation Scenarios

To create multiple scenarios for evaluating the possibility of performance extrapolation across environments, for both algorithms we filtered the collected raw data (i.e. we used different parameterizations of algorithms) as follows. Firstly, we evaluated the localization errors in case measurements from all APs were used as inputs to an algorithm (including the APs that we have not deployed and using both 2.4 and 5 GHz Industrial, Scientific and Medical (ISM) frequency bands). Secondly, we used only measurements from the 2.4 GHz ISM frequency band. Thirdly, we used only measurements from “dedicated” APs, i.e. the six APs we deployed for localization purposes. Furthermore, we filtered measurements from only some dedicated APs to evaluate if the performance degradation due to a removal of particular APs is consistent across environments. Our decision on which AP to include was not driven by the goal of optimizing the APs deployment, but merely to increase the diversity of evaluation scenarios.

However, in case the environments are less similar, the feasible questions that can be addressed, under the assumption that the relative difference in performance results across environments is preserved, are more limited. According to

that limitation, we designed our evaluation scenarios. In the first scenario, we used only six dedicated APs. In the following scenario, we used the APs 1, 2 and 4, which corresponds to a scenario in which two APs are on one side and one AP is in the center of the other side of an environment. In the third scenario, we used the APs that are all on one side of an environment. Finally, in the last two scenarios, we filtered one AP from the corner and from the center of an environment, respectively. The goal is to evaluate the possibility of preserving the relative difference in performance across environments for different scenarios, despite the fact that differences in environments and parameterizations exist.

5.3 Evaluation Approaches

The localization errors in fingerprinting generally have two aspects: spatial and temporal [13]. The temporal one is related to changes of measurements in time domain, which usually causes inconspicuous differences between training and runtime fingerprints. The spatial aspect is related to a particular environment and respective algorithm’s parameterization. We aimed on removing the temporal aspect and focusing on the spatial one, so we do not clearly differentiate between the training and runtime phases of fingerprinting, i.e. we use measurements from the same measurement campaign in both phases. We leveraged two approaches to obtain the split between the training and runtime phases. Firstly, we used first three RSSI measurements from each AP at each location for training, while the fourth measurement was used in the runtime phase. In the reminder of this paper we will refer to this approach as “evaluation approach 1”. Secondly, while a given measurement point was evaluated, measurements from all other points were used for training, and we will refer to this approach as “evaluation approach 2”. Two approaches were selected to increase the number of evaluation scenarios, thus increasing the reliability of our findings.

6. EVALUATION RESULTS

In this section, we present the results of evaluating the possibility of extrapolating the performance of fingerprinting algorithms across environments. Given the observations from Section 3 and Section 4, we classify one environment and algorithms’ parameterization as “similar” (TWIST partial), while the other two are “less similar” to the baseline hospital environment (TWIST and w-iLab.t I).

6.1 Extrapolation in Similar Environments

The distributions of localization errors for the two used algorithms in similar environments are given in Figure 2 and Figure 3 for the evaluation approaches 1 and 2, respectively. As visible in the figures, for both algorithms and for both evaluation approaches the localization errors in similar environments show small statistical differences. The results show that it is possible to give a statement about the performance of fingerprinting algorithms in one environment by using the performance results from another, similar environment. The results also validate the statistically repeatable performance of the used algorithms across similar environments.

Furthermore, we evaluated the similarities in distributions of localization errors across environments using Pearson’s Chi-Squared tests. The results yield that, for all evaluation scenarios, both evaluation approaches, and both algorithms, the hypothesis of localization error distributions across environments being statistically comparable is true with the

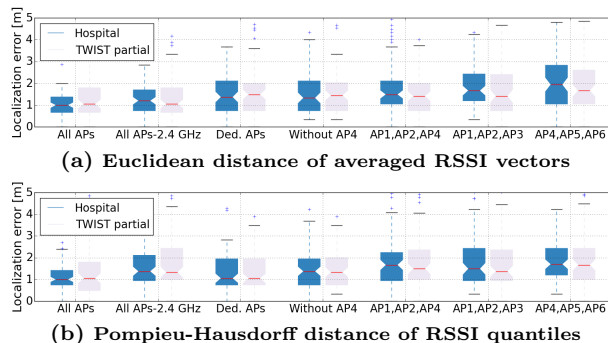


Figure 2: Comparison of localization error distributions for similar environments - evaluation approach 1

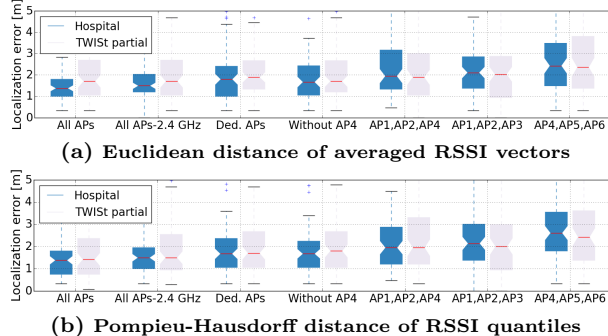


Figure 3: Comparison of localization error distributions for similar environments - evaluation approach 2

probability of more than 95%. Finally, we used the Cohen’s d tests to evaluate the magnitude of the difference of mean localization errors across environments. The result of a Cohen’s d test is Cohen’s d value, which is a scale-free indication of the size of an effect between two observations [14]. As a rule of thumb, Cohen’s d values smaller than 0.2 represent small effect, values smaller than 0.5 represent medium size effect, while higher values than 0.5 represent high effect size. Specifically, in our evaluation Cohen’s d values represent the magnitude of the effect that the change of an environment has on the achieved localization errors. Table 3 gives the Cohen’s values for localization error distributions achieved by the two used algorithms for different evaluation scenarios across two similar environments. As visible in the table, the calculated Cohen’s values are smaller than 0.2, meaning that the change of environments in this case has a small effect on the localization errors. Higher Cohen’s d values are obtained in case all APs in the environments were used in the evaluation (“All APs” and “All APs-2.4 GHz”), since in these scenarios also uncontrollable APs (e.g. visible APs from neighboring buildings) were used as inputs to the algorithms, which resulted in a higher effect size that a change in environments has on the achieved localization errors.

In order to get a clearer view on the possibility of performance extrapolation of fingerprinting algorithms across environments, in Figure 4 we depict the localization error distributions achieved by the used algorithms in different rooms in two similar environments. Furthermore, in Table 4 we present the Cohen’s d test results for the distributions of localization errors per room for different evaluation scenarios across similar environments. Due to lack of space, we present only results obtained by leveraging the evaluation approach 1, and this will be followed through the rest of

Table 3: Cohen’s d test results achieved in similar environments for different evaluation scenarios and for both evaluation approaches (approach 1/approach 2)

All APs	All APs 2.4GHz	Dedicated APs	Without AP4	AP1,AP2,AP4	AP1,AP2,AP3	AP4,AP5,AP6
Euclidean distance of averaged RSSI vectors						
0.45/0.58	0.06/0.39	0.03/0.18	0.09/0.06	0.11/0.12	0.08/0.07	0.09/0.07
Pompiou-Hausdorff distance of RSSI quantiles						
0.16/0.35	0.14/0.34	0.02/0.19	0.06/0.24	0.02/0.02	0.07/0.12	0.06/0.03

Table 4: Cohen’s d test results for the localization errors per room achieved in similar environments for different evaluation scenarios - evaluation approach 1 - (room 1/room 2/room 3/hallway)

Dedicated APs	Without AP4	AP1,AP2,AP4	AP1,AP2,AP3	AP4,AP5,AP6
Euclidean distance of averaged RSSI vectors				
0.11/0.07/0.38/0.09	0.19/0.01/0.29/0.15	0.01/0.11/1.15/0.12	0.22/0.07/0.79/0.09	0.23/0.16/2.79/0.11
Pompiou-Hausdorff distance of RSSI quantiles				
0.17/0.15/0.85/0.03	0.12/0.22/0.60/0.11	0.05/0.09/2.00/0.08	0.16/0.11/1.30/0.19	0.21/0.16/2.91/0.14

the paper, since the results, except for the values of the obtained localization errors, are consistent for both evaluation approaches. The rooms in Figure 4 are labeled with “1”, “2”, “3” and “Hall”, and those are respectively the rooms where APs 1, 2 and 3 are deployed and the hallway (Figure 1a, Figure 1b). It is clear from Figure 4 and Table 4 that the change of environments has a small effect on the achieved localization errors per room, except for the errors achieved in the room labeled with “3”.

The reason for this trend lies in the fact that only 6 measurement points have been defined in the room labeled with “3”. Due to that, no meaningful statistics about the localization errors in this room could be extracted. In other words, it is not our intention and we believe it is not possible to extrapolate the localization errors for a single measurement point from one environment to another. The reason why a single evaluation point is not sufficient lies in the instability of the RF-based indoor localization algorithms and solutions, which is related to an intrinsic randomness in each wireless environment. This means that, even for the same environment and exactly the same measurement point, if a localization estimate is requested twice, these estimates are usually not the same. On the contrary, we believe it is only possible to accurately extrapolate the statistics of errors, and for obtaining a statistic a meaningful number of samples has to be used. This requirement is not fulfilled for the samples in the room labeled with “3”, thus the performance of algorithms is not statistically comparable in this case.

6.2 Extrapolation in Less Similar Environments

In case less similar environments are used for the performance evaluation, the results are depicted in Figure 5. Despite a smaller level of similarity between those environments and the hospital, we aim on evaluating the possibility that the relative difference between the achieved localization errors for different parameterizations of algorithms can be preserved. Note that for this case we do not evaluate scenarios in which all measurements are used, but only scenarios where measurements from the six dedicated APs are leveraged. The reason is that, due to larger sizes of these environments in comparison to the hospital, the number of visible APs is expected to be increased, which could lead to wrong conclusions about the extrapolation possibility.

As visible in Figure 5, the evaluation results show that a relative difference between various parameterizations of algorithms is preserved for the TWIST environment, while for the w-iLab.t I environment that is not the case. More specifically, it is visible in the figure that, for the hospital and the TWIST environment, higher localization errors are achieved in case APs from one side of the environment are used as

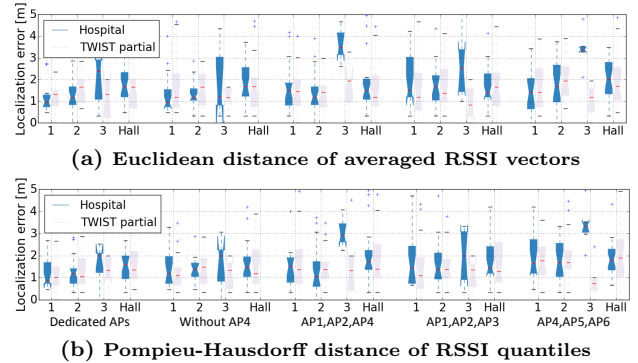


Figure 4: Comparison of localization error distributions per room for similar environments - evaluation approach 1

anchors for location estimation (AP1,AP2,AP3 in figure), in comparison to the scenario when two APs are used on one side and one anchor on the other side of the environment (AP1,AP2,AP4). This observation, however, does not hold for the w-iLab.t I environment, since in this case the localization performance is better when only APs from one side of the environment are used. Similarly, for both algorithms, and for both hospital and TWIST environments, comparable localization errors are achieved in case an AP from the center of an environment is not used for the localization purposes (Without AP4), in comparison to a case when an AP from the corner of an environment is not used (Without AP1). As it can be seen in the figure, in this case the ranking is again not preserved for the w-iLab.t I environment. Similar observation can be made from Table 5, since Cohen’s d values are similar across different evaluation scenarios in case the hospital is compared to the TWIST environment, while they have larger discrepancies in case the hospital is compared with the w-iLab.t I environment.

Clearly, the hospital environment and the two other, less similar environments differ in the outer sizes, as well as in the number of inner walls. However, the hospital and TWIST environments have comparable room sizes, while that is not the case for w-iLab.t I environment. Furthermore, the wall attenuation factor is different for all three environments. This factor is much higher for the TWIST environment, in comparison to the hospital, while for the w-iLab.t I this factor is much smaller than for the hospital environment. A relatively high wall attenuation indicates that a wireless environment has higher spatially distinguishable features, which benefits fingerprinting in general [13]. This is possibly a reason for substantially smaller localization errors in TWIST, in comparison to w-iLab.t I environment, in case a small number of APs is used. In all three environments, the

same number and configuration of APs are used, although their deployment locations and densities are different. Finally, the number and density of measurement points differ, while the number of measurements taken at each point is the same for all three environments.

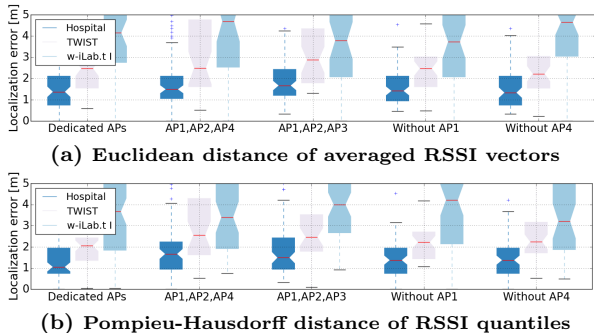


Figure 5: Comparison of localization errors for less similar environments - evaluation approach 1

Table 5: Cohen’s d test results achieved in less similar environments - evaluation approach 1

Ded. APs	AP1,2,4	AP1,2,3	No AP1	No AP4
Euclidean distance of averaged RSSI vectors				
0.68/1.41	0.69/1.04	0.64/1.39	0.79/1.67	0.84/1.32
Pompieu-Hausdorff distance of RSSI quantiles				
0.74/1.56	0.72/1.27	0.77/1.04	0.82/0.47	0.72/1.84

7. CONCLUSION AND FUTURE WORK

In this paper, we postulated a hypothesis for the extrapolation of fingerprinting algorithms’ performance across environments, given that the environments and the parameterizations of fingerprinting algorithms in these environments are similar. We demonstrated that the performance of a set of fingerprinting algorithms in one environment has a small statistical difference to their performance in a similar environment, in terms of the absolute values of localization errors. We have also shown that, even in case environments and parameterizations of algorithms are less similar, the relative performance of fingerprinting algorithms can be preserved. In other words, we demonstrated that the performance extrapolation across environments is a feasible concept, which depends on the similarities among environments.

While we have demonstrated that the performance extrapolation of fingerprinting algorithms across environments is possible, at this point we cannot make any final conclusions on how similar the environments and respective algorithm’s parameterizations have to be. In order to give a reliable answer, insights from additional environments and algorithms are necessary. Our future work will be oriented toward this goal, but we also seek support from the community. The envisioned contribution from the community is twofolds. Firstly, the interested parties can contribute by extending the available datasets with data from additional environments. These data-traces can be provided and used in a simple way through a set of publicly accessible web-based services. Secondly, the interested parties can contribute by evaluating their fingerprinting algorithms using the offered data-traces and provided evaluation services, as described in [8]. By increasing the number of environments and evaluation results, more insights in the correlation between the similarity of environments and parameterizations of algorithms, and the similarity between the achieved performance of such algorithms will be gained. These insights

will serve to evaluate with higher confidence our hypothesis about the possibility of extrapolating the performance of fingerprinting algorithms across environments. Furthermore, it will serve to assess the change in the performance of fingerprinting algorithms due to changes in some of the detected important factors for characterizing the similarities between two environments and parameterizations of algorithms.

In this paper, we have focused on RSSI-based fingerprinting, while future work will also be oriented toward capturing other environment parameters that are influencing the performance similarity of different localization approaches. Our aim is to identify a smaller number of environmental similarity parameters that have a broader impact on the extrapolability of different localization approaches across environments. One clear drawback of our approach is that, at the moment, we have based the assessment of the similarity in a manual way, through the physical shape and propagation characteristics of different environments. To scale the approach we will resort to automatizing it. We are considering to use the SWAT (Stanford Wireless Analysis Tool) [15] for collecting low-level wireless network measurements and using them for automatized quantification of the similarity metrics of different environments.

8. ACKNOWLEDGMENTS

This work has been partially funded by the EU Commission (FP7-ICT-FIRE) within the project EVARILLOS (grant No. 317989). The author Filip Lemic was partially supported by the DAAD (German Academic Exchange Service).

9. ADDITIONAL AUTHORS

Tom Van Haute (Ghent University, tom.vanhaute@intec.ugent.be)
Eli De Poorter (Ghent University, eli.depoorter@intec.ugent.be)

10. REFERENCES

- [1] P. Bahl *et al.*, “RADAR: An In-building RF-based User Location and Tracking System,” in *INFOCOM’00*, 2000.
- [2] D. Lymberopoulos *et al.*, “A Realistic Evaluation and Comparison of Indoor Location Technologies: Experiences and Lessons Learned,” in *IPSN’15*, 2015.
- [3] F. Lemic *et al.*, “Experimental Evaluation of RF-based Indoor Localization Algorithms Under RF Interference,” in *ICL-GNSS’15*, 2015.
- [4] A. Bavier *et al.*, “In VINI Veritas: Realistic and Controlled Network Experimentation,” in *SIGCOMM’06*, ACM, 2006.
- [5] J. Hernández-Orallo *et al.*, “On More Realistic Environment Distributions for Defining, Evaluating and Developing Intelligence,” in *Artificial General Intelligence*, Springer, 2011.
- [6] S. Che *et al.*, “BenchFriend: Correlating the Performance of GPU Benchmarks,” *High Performance Computing Apps*, 2013.
- [7] S. Phansalkar, *Measuring Program Similarity for Efficient Benchmarking and Performance Analysis of Computer Systems*. ProQuest, 2007.
- [8] F. Lemic *et al.*, “Web-based Platform for Evaluation of RF-based Indoor Localization Algorithms,” in *ANLN’15*, 2015.
- [9] F. Lemic *et al.*, “Infrastructure for Benchmarking RF-based Indoor Localization under Controlled Interference,” in *UPINLBS’14*, 2014.
- [10] E. Damosso *et al.*, *COST Action 231: Digital Mobile Radio Towards Future Generation Systems*. EUR Series, 1999.
- [11] G. Caso *et al.*, “On the Applicability of Multi Wall Multi Floor Propagation Models to WiFi Fingerprinting Indoor Positioning,” in *FABULOUS’15*, 2015.
- [12] F. Lemic *et al.*, “Experimental Decomposition of the Performance of Fingerprinting-based Localization Algorithms,” in *IPIN’14*, 2014.
- [13] V. Honkavirta *et al.*, “A Comparative Survey of WLAN Location Fingerprinting Methods,” in *WPNC’09*, IEEE, 2009.
- [14] J. Cohen, “A Power Primer,” *Psychological bulletin*, 1992.
- [15] K. Srinivasan *et al.*, “SWAT: Enabling Wireless Network Measurements,” in *ACM Sensys’08*, ACM, 2008.