Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates

Joaquín Torres-Sospedra*, Philipp Richter†, Germán Mendoza-Silva*, Elena Simona Lohan† and Joaquín Huerta*
*Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain
†Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland
jtorres@uji.es - philipp.richter@tut.fi - gmendoza@uji.es - elena-simona.lohan@tut.fi - huerta@uji.es

***Abstract*—Fingerprinting is widely used for indoor positioning, where pattern matching techniques are usually applied to signals from APs or Beacons. However, the real-time monitoring of the emitters is not an easy task in most cases. When an alteration in the emitters is not detected or properly fixed, it might have a severe impact in the accuracy of the indoor positioning algorithm. Simple but common alterations are energy failure, emitter replacement, wrong emitter placement after maintenance and AP displacement. This paper explores how the AP alteration might be automatically detected by computing the average of the RSS distance to the best match over multiple operational points. The experimental setup consider one simulated and two real scenarios to validate the proposed metric for detecting AP alternation. The results show that it is possible to detect AP alteration when it has a considerable impact in the IPS accuracy.**

I. Introduction

There is no doubt that indoor positioning is a hot research topic since it is an important source of revenues [1] and its application covers many application domains: health, tracking, remote monitoring, advertising, safety, emergency rescue management, among many others.

Among all the existing alternatives, RSS-based techniques using Wi-Fi and/or BLE are of special relevance. Although other technologies, such as Inertial Measurement Unit/Pedestrian Dead Reckoning [2], ultrasounds [3], [4] or Ultra Wide Band [5], [6], might provide more accurate positioning, RSS-based positioning is compatible with the commercial off-the-shell smartphones, which fosters simple and inexpensive software solutions for indoor positioning and navigation.

Wi-Fi fingerprinting relies on a pre-collected set of RSS measurements at different locations within the survey area, that ideally cover the whole area of interest. These RSS measurements and the associated locations are usually stored and maintained in a database, also called radio map. The localization then takes place in an operational phase during which the estimated location is computed using the positions of the most similar fingerprint or set of fingerprints.

One of the biggest drawbacks of fingerprint-based positioning is generating the radio map and maintaining it up-to-date over the time as the environment or the network infrastructure is subject to change – changes that in turn alter the spatio-temporal distribution of the RSSs. Those changes may occur because of temporal or permanent failures, maintenance tasks, Wi-Fi coverage optimization, human-errors or, even, vandalism. Four typical changes that we have detected are:

- Emitter(s) stop broadcasting because of temporal malfunction, energy loss/battery drain, or permanent removal,
- Emitter(s) are replaced because of a failure or upgrade,
- Emitter(s) are swapped after maintenance and
- Emitter(s) are slightly moved or displaced.

These alterations are rather long lasting compared with ephemeral deviations from the radio map, such as noise [7] or minor changes in the environment caused by opening/ closing of doors or the presence of people [8]. Changes in the network infrastructure degrade the positioning performance much more than the ones due to noise or minor changes in the RSS distribution [9], e.g., changing the orientation of an AP. Smaller deviations from the initial RSS distribution are usually compensated by the redundancy provided by a large number of APs, whereas larger changes may cause significant positioning errors [10]. In order to retain the positioning performance, events that cause a significant mismatch between the previously sampled RSS distribution and the current RSS distribution need to be detected and, if permanent, reflected by the radio map.

Current trends to mitigate the costs of generating the radio map are to use propagation loss models [11], [12], regression methods [13], [14] or other advanced alternatives [15]. Attempts to reduce the costs of updating the database are primarily crowdsource-based [16]–[18]. Although the database can be maintained with crowdsourced data, the location labels are often imprecise, which decreases the radio map accuracy.

Work related to the detection of RSS alterations comprises studies about attacks on, or faults in, fingerprinting localization systems and studies about RSS outlier detection based on temporal data. In [19], the minimum distance between the RSS observation vector and the RSS database entries is used to detect signal attenuation and amplification attacks. Laoudias et al. [20] build upon that approach to detect AP faults. They extended it with a robustified position estimator. The authors of [21] address access point faults by validating the observed RSSs against the fingerprint that yielded the smallest distance plus/minus its standard deviation; this method presumes several RSS per fingerprinting in the radio map . He et al. [14] also rely on the difference between measured RSS and fingerprints in the database to indicate an alteration of the measured RSS, but they analyse in addition the dispersion of estimated positions from these RSS to deduce a possible RSS alteration. The authors of [12] compare the effect of different attacks and faults on the positioning accuracy. Other works focus on detecting a deviation of the RSS by time-domain processing [22]–[24] presented an outlier detector based on robust statistics derived from the RSS in a time window.

This study presents a thorough analysis of several types RSS alterations based on the statistics of two indicators: the average distance between RSS test measurements and RSSs of the database and the resulting position estimates; the position estimates are also exploited to infer their spatial proportions. These indicators are analysed for three different databases and three frequently encountered fault scenarios. We pay special attention to the effect of virtual networks versus physical APs.

II. Materials and Methods

This section introduces the metric and the statistics derived from it that we use to infer information about alterations of APs. Subsequently, we describe the experimental setup used to assess their effectiveness for certain scenarios. To allow the reproducibility and further extension of this paper's experiments, the software and databases generated are publicly available at [25] in order to promote reproducible research [26].

A. The indoor positioning algorithm: k-NN

Despite the diversity of Wi-Fi fingerprinting methods, the k-NN algorithm [27] is the core of some of them [28]–[30]. Consider a total number of L APs and a radio map consisting of M reference points. Given an operational fingerprint, $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_L]$ whose collection position is unknown, this pattern-matching algorithm calculates its distance in the feature space (or RSS space) to a set of fingerprints previously recorded and stored in a reference dataset (radio map), $X = \{\mathbf{x}_m\}_{m=1}^{M}$, where $\mathbf{x}_m = [x_{1,m}, \ldots, x_{L,m}]$: $d_m = \|\tilde{\mathbf{x}} - \mathbf{x}_m\|$.

Then, the position is estimated by using the position associated with the most similar reference fingerprint, i.e., the fingerprint that yielded the smallest RSS distance (distance to the best/closest match). This corresponds to the k-NN with $k = 1$. Let $\mathbf{d}$ denote the M distances derived from X, then the smallest distance is simply $\hat{d} = \min(\mathbf{d})$. With the index of the corresponding reference fingerprint being $\hat{m} = \arg\min_m(\mathbf{d})$, we denote the position estimate by $\tilde{\mathbf{p}} = \mathbf{p}_{\hat{m}}$.

For a set of test fingerprints, the outcomes of the k-NN algorithm are the smallest RSS distances and the resulting position estimates. The following two subsections detail their use as indicators of alterations of the RSS distribution.

1) Smallest distance in feature space: To infer changes of the APs we use basic descriptive statistics of the RSS distance for different sets of test data. As the test sets either contain or do not contain AP alterations, a comparison of the statistics of the smallest distances in feature space may reveal information about AP alterations. It is the main indicator of AP changes used in this study.

2) Position estimates from smallest RSS distances: Similarly, we compute the statistics of the resulting position estimates of the different test data sets to evaluate its usefulness to indicate AP changes. A comparison of the statistics between the fault free and a faulty case may reveal AP alterations.

The position estimates also comprise spatial information. To exploit that information we approximate the spatial distribution of the position estimates with histograms. To increase the significance regarding changes of APs, we compute the gradient of the histograms and represent them with truth values (logical values), which is enough to indicate if an AP was altered or not. We can then compare the matrix of truth values of data sets which contain AP changes with those that do not.

For this spatial analysis, we use $k = \{3, 5\}$, instead of $k = 1$ for the NN to increase the accuracy but also the spatial spread of the position estimates. In case of the `libdb`, the radio map data for each floor was processed separately, however, for each floor the test data comprising both floors was used.

B. Databases

To study the impact of changes in the Wi-Fi network on the distance in the RSS space, we selected three data sets. First, a completely controlled data set whose RSSs were simulated [31]. This dataset facilitates our analysis as all its properties are well known. The second data is from a real, but relatively small, environment [32] and the third data set is from a medium-scale, environment [33]. Properties of the latter two data sets and the corresponding environments are partially known.

`simdb` is a simulated dataset that is generated with a simple path loss model consisting of a fix transmit power, free-space path loss and a zero mean, normally distributed random noise with a 2 dBm standard deviation. The fingerprints are on a 1 m grid in a 50×20 meter environment, thus, the distribution of fingerprints is dense. A fingerprint at a certain position consists of the RSS of eight AP that are regularly and symmetrically distributed along the top and bottom edges of the environment.

`libdb` is a database collected on two floors of a university library during 15 cumulative months. The area covered by that data set is about 15×10 meter and the average spatial distance between fingerprints is about two meter. This data contains information from virtual Wi-Fi networks, i.e., different MAC addresses from the same physical AP in possibly different spectra. For more details about this dataset we refer to [10]. For this work, we use only the first ten months of the data.

`mandb` is a database that covers the corridors of a university department of about 50×36 meter. The fingerprints are on a 1.5 m grid [34]. The locations of 11 APs are known.

In order to increase the reliability of the results and to provide additional statistical information, we created multiple subsets for each dataset: 100 for `simdb`, 60 for `libdb` and 110 for `mandb`. The number of subsets has been set according the features of each database as detailed in [25].

C. Manipulation Scenarios

Changes of the RSS distribution due to alterations in the AP deployment are simulated by manipulating the test data of the data sets. That is, we assume an unaltered fingerprint radio map, the nominal case, but altered RSS measurements are injected during the operational phase.

The following manipulations correspond to the most common scenarios detected in our facilities in a period of five years.

`Nominal` This scenario corresponds to the normal scenario with no AP alteration. It corresponds to the simple run of the NN algorithm for each subset of the database.

`Remove antenna` This scenario simulates that a particular antenna has been switched off during the operational phase because it does not work or it has been replaced. Training data are kept unaltered and the RSS values for the selected antenna are filled with non-detected values (+100).

`Remove antennas randomly` This scenario simulates that a set of n random antennas have been switched off during the operational phase. The antennas are randomly selected for each subset of the database.

`Swap antennas randomly` This scenario simulates that a set of n random antennas have been swapped during the operational phase. The antennas and the new locations are randomly selected for each subset of the database.

III. Experiments

This section introduces the results of the experiments on different manipulation scenarios for each of the three databases.

A. Results with `simdb`

First, the effect of disconnecting antennas in the distance in the feature space was evaluated. Figure 1(a) shows the boxplot of the distance in the RSS space to the best match for all the evaluation points and all subsets (100×1000 values), whereas Figure 1(b) shows the boxplot with the average distance of all the evaluation points per subset (100 values).

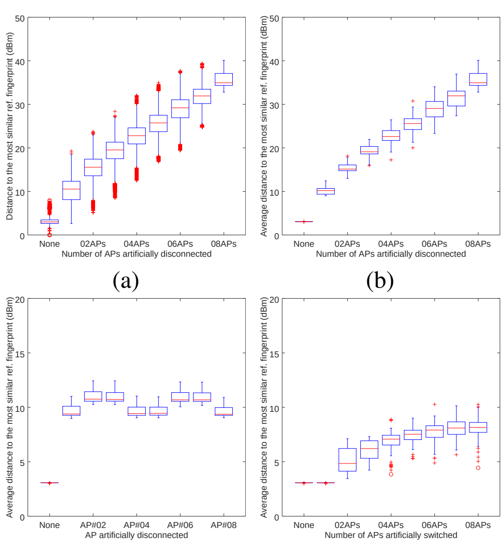


Fig. 1. Impact of basic AP manipulations in the distance to the closest match in the RSS space: (a,b) removing n antennas, (c) removing a particular AP, and (d) swapping n antennas randomly.

For `simdb` it is clear that the higher the number of removed antennas, the higher the distance in the RSS space to the best match. Moreover, the average over all evaluation points is more significant and robust since its statistical spread is much lower.

Second, the effect of disabling a particular antenna in the distance in the feature space was evaluated. Figure 1(c) shows the average distance in the RSS space for this analysis. Although all the APs are important, it seems that the ones located near the scenario corners ($AP_{1,4,5,8}$) have slightly less impact in fingerprinting. Next, we divided the area in eight zones, each corresponding to the area in the vicinity of one of the eight APs, and computed the average of the smallest RSS distance over all data falling in each zone. Figure 2(a) shows the effect of disconnecting a particular AP for Zone 1, the zone closest to AP_1. The largest RSS distance is provided after removing AP_1, followed by the RSS distances of removing AP_2 and AP_5, which are direct neighbours of AP_1. The same holds if we look at Zone 7, for which the largest RSS distance is obtained after removing AP_7, followed by AP_6, AP_8 and AP_3. This shows that positions estimates yielded from the smallest RSS distance comprise useful information about the AP alterations.

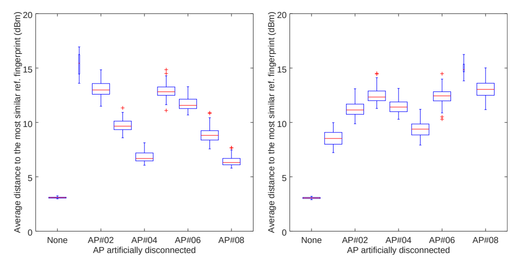


Fig. 2. Impact of removing all 8 APs on the distance to the closest match in the RSS space for Zone 1, (a), and Zone 7, (b). The positions estimated from the smallest RSS distance have been used to determine the zone.

Third, the effect of randomly swapping some antennas was also evaluated, see Figure 1(d). Similarly to the case where the APs were disconnected, the more antennas are swapped, the higher the impact in the average distance to the closest match. As it happened with the AP removal, it is expected that the effect of switching antennas will depend on the location and APs switched. But that study is out of this paper scope.

To summarize, detecting an AP change cannot be done with the RSS distance of a single fingerprint, but the averaged distance in the RSS space over a larger set of evaluation points might be a good indicator of AP manipulations. Additionally, the position estimates comprise information about the AP manipulations, too; which confirms the notion that some APs might be more relevant in certain areas than other APs.

B. Results with `libdb`

This database comprises six RSS measurements per reference and evaluation point. By selecting the six measurements per reference point but just one measurement per evaluation point for 10 months, we formed 60 subsets (10 months×6 fingerprints). The results are averages over these 60 subsets. Prior the experiments, a manual AP selection was performed to get the representative APs in the environment belonging to the official university Wi-Fi networks (a total of 52 APs).

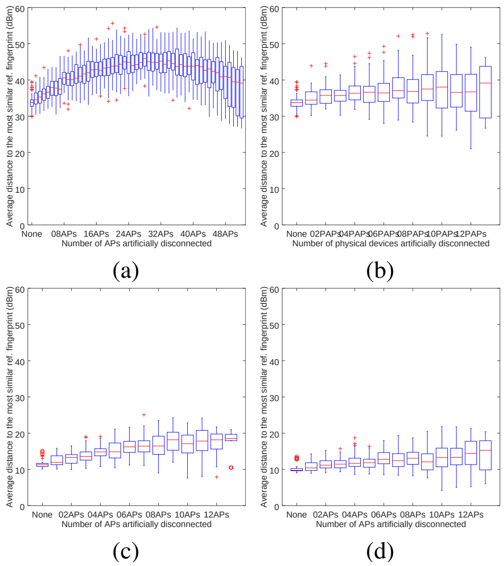


Fig. 3. Impact of randomly removing n APs on the average distance to the closest match in the RSS space: (a) considering the 52 APs independently, (b) considering that the 52 APs were emitted by just 13 physical devices ($13 \times 4 = 52$), (c,d) considering 13 APs by merging the 2 RSS from the 2.4 GHz and 5.2 GHz bands respectively.

Figure 3(a) shows the effect of disconnecting n antennas at the operational stage by using the reduced set of 52 APs. In contrast to `simdb`, the RSS distance metric increases until the 50% of APs are removed. After that, the metric starts to decrease. This effect might be due to the presence of virtual networks in the Wi-Fi deployment. The 52 APs are not independent because 13 devices were broadcasting 4 different SSIDs, 2 in the 2.4 GHz band and 2 in the 5.2 GHz band. The redundancy from the remaining virtual APs of the same AP compensates the fault case. Therefore, the impairment on the RSS distance due to such faults is small. The variability is high in all cases and increases with increasing number of disabled APs. To consider the presence of virtual APs, Figure 3(b) shows the effect of disconnecting physical devices instead (disconnecting a physical device means removing its four broadcast networks and the associated RSS data). Although the relation between RSS distance and the alteration is slightly clearer in Figure 3(b), the high variability persists.

However, we detected that one of the two SSIDs in the same band was not detected sometimes. Although this was expected for weak RSS values near the sensitivity threshold, it was also detected for stronger values in a few cases. Thus, the analysis was extended to fuse all the SSIDs emitted in the same band by a device. Figure 3(c,d) shows the results when the RSSs in the 2.4 GHz and 5.2 GHz band are independently analysed. It suggests that the APs in the 2.4 GHz are more informative regarding the detection of changes of the RSS distribution. The lower frequency (the lower attenuation) allows signals of the 2.4 GHz networks to cover a larger range that leads to more measurements compared to the 5.2 GHz ones (see Figure 4). The number of measurements in the 5.2 GHz band is lower, which provides less information about the AP alterations.

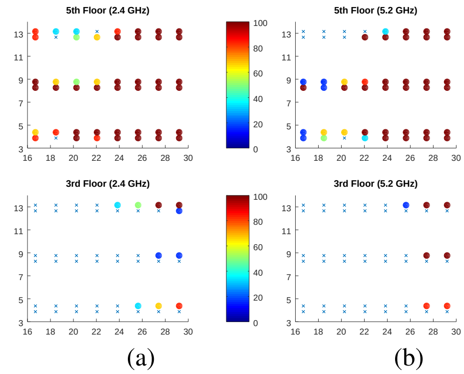


Fig. 4. RSS measurements (in dBm shifted to positive axis) from AP_6 on two floors, separated for the 2.4 GHz and 5.2 GHz band. Top plots show the upper floor and bottom plots show the bottom floor of the environment.

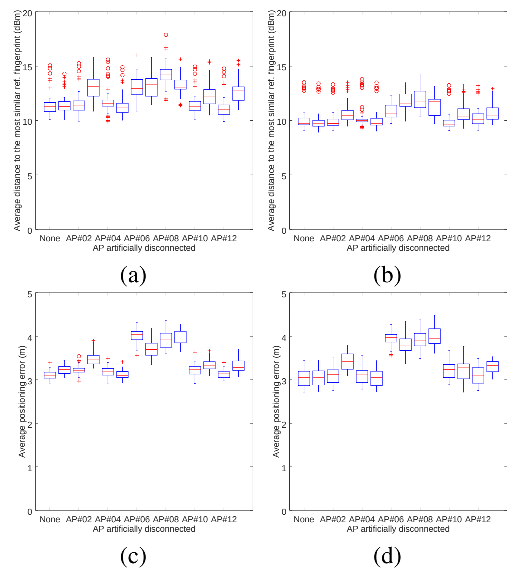


Fig. 5. Impact of disconnecting a particular AP in the distance to the closest match in the RSS space and positioning error. Reduced data sets considering fused RSSI values in the 2.4 GHz band (a,c) and 5.2 GHz band (b,d).

Figure 5 compares the averaged RSS distance to the closest match and the average of the resulting position error over the data subsets, Figure 5(a,c) correspond to the 2.4 GHz band, whereas Figure 5(b,d) correspond to the 5.2 GHz band. Figure 5(a,b) depict the distance in RSS space and (c,d) show the positioning error. The effect of an AP alteration is more distinctive in the positioning error. It can also be seen that the $APs_{\{6,7,8,9\}}$ impact the RSS distance metric and the positioning error more than the other APs and can be considered more more relevant for the IPS than the remaining APs. The $APs_{\{3,11,13\}}$ are semi-relevant, because they show only a slight difference in both metrics compared the remaining APs that do not show a significant alteration. It is worth mentioning that around 50 % of APs are not relevant, which could explain the high variability in the plots shown in Figure 3. This methodology of artificially removing single APs is useful to determine the relevance of an AP for the IPS.

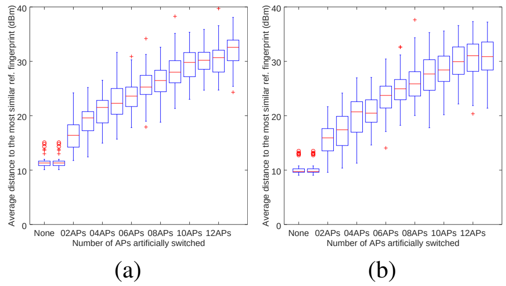


Fig. 6. Impact of swapping sets of APs in the distance to the closest match in the RSS space. Reduced data sets considering RSS values from APs in the 2.4 GHz band (a) and from APs in the 5.2 GHz band separately (b).

The averaged RSS distance metric for the AP swapping scenario is presented in Figure 6. The manipulation of AP was done separately for networks in the 2.4 GHz (a) and in the 5.2 GHz (b) bands. The results confirm that the effect swapping the location of APs on the distance in the RSS space is significant. It also seems that the data from the 2.4 GHz band APs suit better to detect changes in the environment.

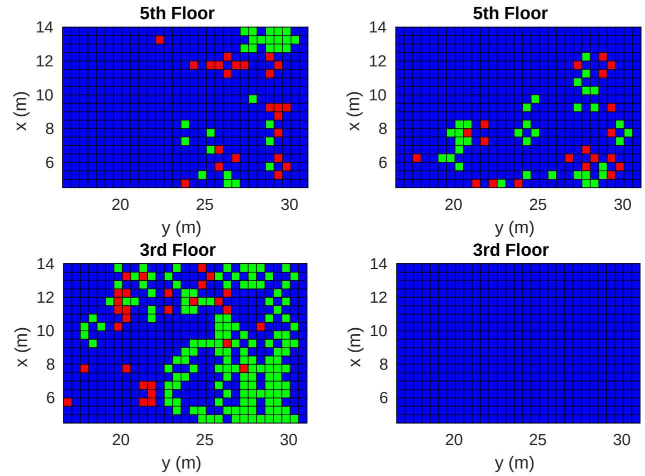
Finally, we evaluate the spatial distribution of the position estimates as an indicator for AP alterations. The method described in section II-A2 is applied to the scenario of removing particular APs, considering only the 2.4 GHz band. The result for AP_6 and AP_{10} is depicted in Figure 7. The blue cells indicate that, compared to the nominal case, no significant change of position estimates took place. The red indicates a decrease of the amount of position estimates in the corresponding cells and green indicates an increase of position estimates. First of all, AP_6 affects more cells than AP_{10}. That is conclusive with Figure 5 from which we concluded that AP_6 is more relevant than AP_{10}. Furthermore, most changes are located on the right side of the figures showing alteration of AP_6, specially in the third floor. This observation coincides with Figure 4 which strongly suggests that AP_6 is located on the fifth floor near the right border.

Alterations of some APs hardly affect the RSS distance. These APs are irrelevant for the IPS, because neither of them cause an increase the positioning error. The distance in the RSS space to the closest match cannot cope with virtual APs due to the inherent redundancy and with APs of low relevance. To use this metric for the detection of AP faults, in general, suitable subsets of APs need to be considered, but additional knowledge about the network infrastructure is usually required to find those subsets. The spatial analysis of the positioning estimates not only indicates AP manipulations but also indicates the area that is affected by a particular AP.

C. Results with `mandb`

After the initial assessment with a path-loss based database and the validation with a real data set collected in one of our facilities, we used an external data set for final validation.

The Mannheim data set includes 110 fingerprints per reference and evaluation point. In order to consider the fingerprint heterogeneity, a total of 110 data sets were generated. The i-th data set contains the i-th fingerprint of each evaluation point, whereas 10 fingerprints per reference point have been randomly extracted. However, for the analysis described in section II-A2 we used a single fingerprint of the radio map but all 110 RSS measurements of the test data.



AP_6 AP_{10}

Fig. 7. Impact of disabling a particular AP on the relative distribution of position estimates. Plots on the left hand side show such a distribution for AP_6 and the plots on the left hand side depicts it for a less relevant AP, namely AP_{10}, both for the 2.4 GHz band.

We have considered two versions of the database according to the APs used for fingerprinting: a first one with all the detected APs (28 APs); a second one with the 12 APs whose location was provided in the data set. However, one of the APs in the later case was never detected and, therefore, the second version only includes 11 APs (see Figure 8), being all of them in the environment nearby and 10 of them on the same floor; AP_7 is not shown here, it is located on an adjacent floor at a similar location as AP_6.

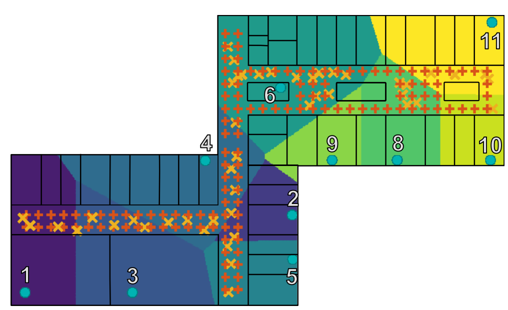
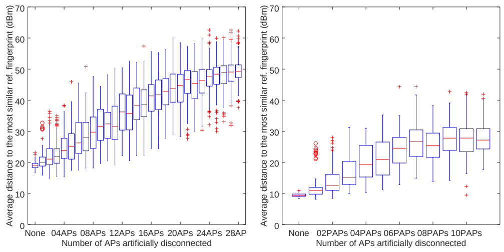


Fig. 8. Mannheim floormap. The blue circles are the location of APs and the coloured polygons represent the zones.

First, the effect of disconnecting n antennas randomly is depicted in Figure 9 for the full data set with 28 APs (a) and the reduced version with only 11 APs (b). The trend is clear, the higher the number of removed APs, the larger the impact in the average distance in the RSS space, specially in the reduced data set (11 APs). The variability seems to be higher for the full data set than the reduced one, making a detection of AP

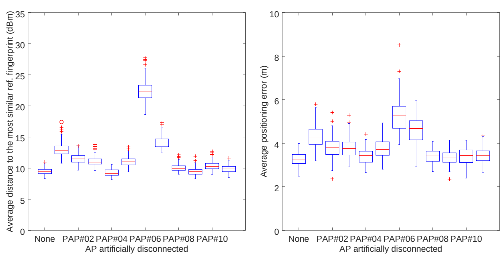


(a) (b)

Fig. 9. Impact of disconnecting n random APs in the distance in the RSS space to the closest match and positioning error.

alterations more difficult. This difference is probably due to the high number of irrelevant APs (up to 17) located far from the scenario in the former data set.

The analysis of disabling a particular AP resembles the results obtained with `libdb`. Figure 10 shows that alterations of relevant APs are observable in the distance in the RSS space to the closest match and also in the positioning error, whereas the RSS distance metric and the positioning error of less relevant APs do not deviate significantly from the corresponding quantities of the nominal case. Accordingly, $AP_{\{1,2,3,5,6,7\}}$ are considered relevant.



(a) (b)

Fig. 10. Impact of disconnecting a particular AP in the distance in the RSS space to the closest match and positioning error.

Additionally, we analysed how the position estimates distribute spatially compared to the nominal case, which is shown for $AP_{1,6,8,11}$ in Figure 11. This analysis is similar to the one shown in Figure 7 for `libdb`. In the figure, the white cells contain no position estimates. A comparison with Figure 8 reveals in the two top plots a decrease of position estimates in cells close to the antennas location. Bottom left plot shows that there are many green and red cells mostly spread in the vertical and the upper horizontal corridors, actually nearby AP_6. Bottom right plot shows fewer green and red cells compared to the other cases, most of them in the right hand side corridor where AP_8 is indeed located. These graphics indicate the region(s) that are affected by the corresponding alterations and additionally enable a rough localization of the AP that caused the relative change of position estimates.

In contrast to the `simdb` and `libdb`, the AP distribution in `mandb` creates irregular zones of influence. Moreover, the distribution of operational fingerprints is not equally distributed over all the scenario, they are only located in the corridors,

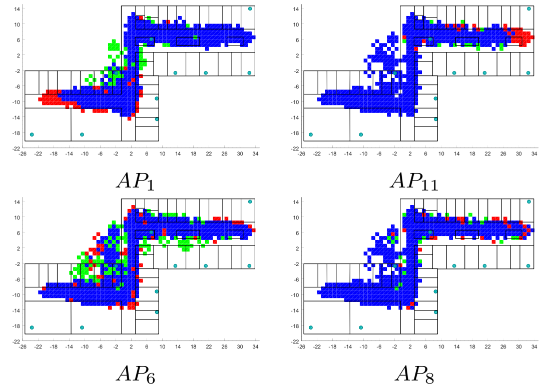


Fig. 11. Impact of disabling a particular AP on the relative distribution of position estimates when AP_1, AP_{11}, AP_6 or AP_8 are removed.

but not in the adjacent rooms and almost 33 % of them are placed near AP_6. Which could explain why AP_6 and AP_7 are of special relevance in this data set.

To better understand the relevance of the 11 APs whose location is well-known, the analysis was performed also in all the zones of the scenario (see Figure 12). The ith-zone of this scenario corresponds to the area to which the nearest AP is AP_i, see Figure 8. Zone 7 is not included because AP_7 is located in a different floor close to AP_6. According to the results shown in the figure, $AP_{1,2,3,5}$ are of special relevance in the zones located in the bottom-left horizontal corridor (Zones 1 to 5). Moreover, AP_6 is relevant in the vertical corridor and the top-right horizontal corridor (Zones 2 and 4 to 11), whereas AP_7 seems to be only relevant in the top-right horizontal corridor. Manipulating the other APs shows almost no impact on the averaged RSS distance metric in any zone. Even in a few zones, the averaged RSS distance to the best match is decreased after removing AP_4.

The resulting averaged RSS distance from the experiments with the scenario `Remove antenna` suggest that the relevance of APs differs, and that it is location dependent. The corresponding positioning errors (Figure 5(c,d) and Figure 10(b)) indicate further that APs which we consider relevant have a strong effect on the positioning performance, whereas the APs which we consider of low relevance have only a small effect on the positioning accuracy.

Figure 13 confirms this, as it shows a moderate-high correlation between the positioning error and the distance to the best match in the RSS space according to two correlation analyses, Pearson and Spearman. It depicts the relative average RSS distance to the best match against the relative average positioning error, computed with 1-NN on all the subsets of the reduced versions of `libdb` (60 subsets, 13 2GHz APs) and `mandb` (110 subsets, 11 APs) for the `Nominal`, `Remove antennas randomly` and `Swap antennas randomly` scenarios. Both metrics show relative values with respect to the nominal case for each database subset. The linear regression curve has a slope of 1.3. Indeed, also previous works have shown that APs can have a different relevance for positioning and that it depends on the AP location. This fact is used to reduce the size

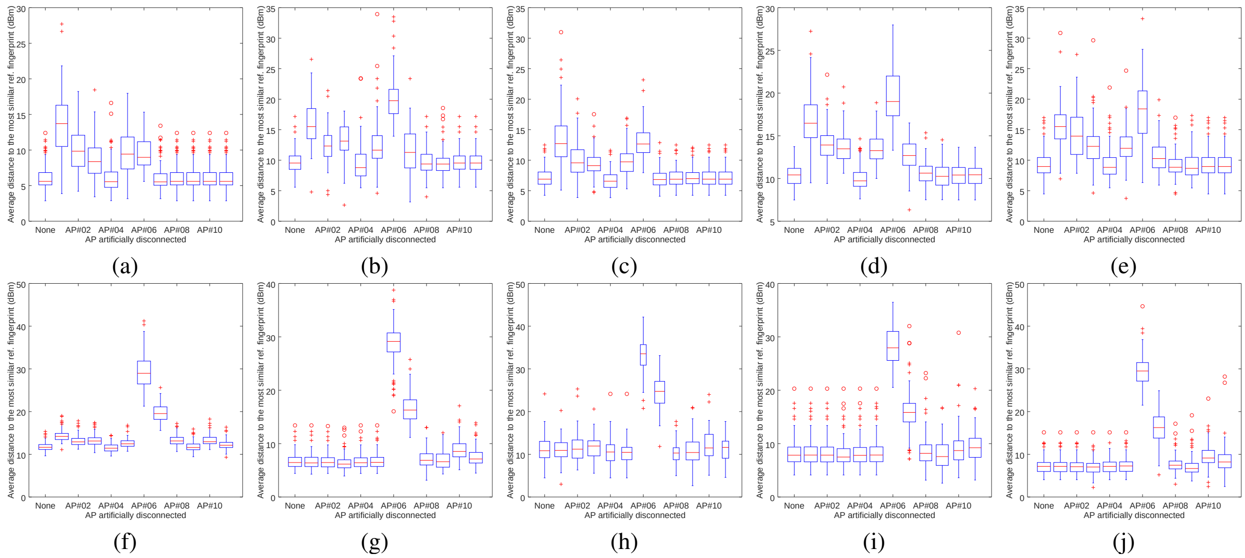


Fig. 12. Impact of basic AP manipulations in the distance to the closest match in the RSS space in each zone: (a) Zone 1; (b) Zone 2; (c) Zone 3; (d) Zone 4; (e) Zone 5; (f) Zone 6; (g) Zone 8; (h) Zone 9; (i) Zone 10; (j) Zone 11. Different zones reveal different relevant APs.

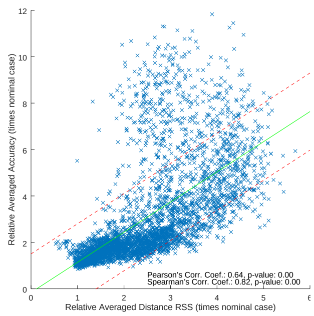


Fig. 13. Relation between relative averaged positioning error and relative distance in the RSS distance. Correlation coefficients are also shown.

of radio maps [35]–[38], to refine the positioning method [37], [39] or to partition the environment [40]. However, these studies apply typically more complex algorithms.

The general observations from the `mandb` agree with those from `libdb`. The RSS distance metric averaged over larger set of evaluation points is a good indicator of changes in the AP deployment if the APs are relevant. Additional information or processing is necessary to detect the alterations of less relevant APs; e.g., zone-based analysis, which could be very useful in very large environments. The spatial analysis of the positioning estimates reflects the results obtained through the averaged RSS distance and can additionally help to locate the altered AP. The number of affected cells compared with its total number may also be exploited to determine changes in the AP deployment.

IV. Conclusions

This paper presented a thorough analysis of common AP alterations based on the averaged distance in the RSS space of the operational fingerprints to the best match. The visual inspection of the statistics derived from that metric and the change of distribution of the resulting position estimates are simple and useful indicators of relevant alterations in the AP distribution.

The spatial distribution of the resulting position estimates can also provide insights about the area, or areas, affected by the alteration of the AP distribution. However, the reliability of the discussed indicators decrease if not enough evaluation points are available. The reliability of the distribution of the position estimates also decreases if the evaluation points do not cover sufficiently the area of interest.

If the objective is the detection of alterations of APs with less impact, of virtual APs or of changes in a specific frequency band, further knowledge about the network deployment and/or the environment is required. The removal of individual APs from the database while analysing the averaged RSS distance metric is as well useful to determine relevant APs. The results also suggest that less relevant APs can be safely removed from the database without scarifying much the positioning accuracy. It is not worth the effort of rebuilding a radio map after an AP alteration that does not significantly impair the IPS accuracy.

Finally, the zone-based analysis might be more useful to detect changes in the APs located in the nearby zone. This would be specially helpful when the operational data come from crowdsourcing and the operational samples are not equally distributed over the operational area.

Acknowledgment

J. Torres-Sospedra and P. Richter have equally contributed to this work. P. Richter would like to thank Mohammad Mehdi for the fruitful discussion. This research collaboration was co-funded by the Academy of Finland (project 313039, http://www.insure-project.org), Universitat Jaume I (grant PRE-DOC/2016/55) and European Commission (GEO-C project, H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http://www.geo-c.eu/).

References

[1] "Markets and markets: Indoor location market worth 40.99 billion usd by 2022," https://www.marketsandmarkets.com/Market-Reports/indoor-positioning-navigation-ipin-market-989.html.

[2] L. Díez, A. Bahillo, S. Bataineh, A. Masegosa, and A. Perallos, "Enhancing improved heuristic drift elimination for step-and-heading based pedestrian dead-reckoning systems," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2016, aug 2016, pp. 4415–4418.

[3] F. Seco, A. R. Jiménez, and F. Zampella, "Fine-Grained Acoustic Positioning with Compensation of CDMA Interference," in *IEEE Int. Conf. on Industrial Technology (ICIT)*, Seville, 2015, pp. 3418–3423.

[4] T. Aguilera, F. J. Álvarez, D. Gualda, J. M. Villadangos, A. Hernández, and J. Ureña, "Multipath compensation algorithm for TDMA-based ultrasonic local positioning systems," *IEEE Transactions on Instrumentation and Measurement*, vol. PP, no. 99, pp. 1–8, 2018.

[5] E. García, P. Poudereux, A. Hernández, J. Ureña, and D. Gualda, "A robust uwb indoor positioning system for highly complex environments," in *2015 IEEE International Conference on Industrial Technology (ICIT)*, Mar. 2015, pp. 3386–3391.

[6] A. R. Jiménez and F. Seco, "Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 2106–2117, 2017.

[7] Y. Gu, M. Chen, F. Ren, and J. Li, "Hed: Handling environmental dynamics in indoor WiFi fingerprint localization," in *2016 IEEE Wireless Communications and Networking Conference*, Apr. 2016, pp. 1–6.

[8] S. Garcia-Villalonga and A. Perez-Navarro, "Influence of human absorption of Wi-Fi signal in indoor positioning with Wi-Fi fingerprinting," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct. 2015, pp. 1–10.

[9] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, Sep. 2014.

[10] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, 2018.

[11] M. U. Ali, S. Hur, and Y. Park, "LOCALI: Calibration-free systematic localization approach for indoor positioning," *Sensors (Switzerland)*, vol. 17, no. 6, pp. 1–16, 2017.

[12] P. Richter, M. Valkama, and E. S. Lohan, "Attack tolerance of RSS-based fingerprinting," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018.

[13] P. Richter and M. Toledano-Ayala, "Revisiting Gaussian process regression modeling for localization in wireless sensor networks," *Sensors*, vol. 15, no. 9, pp. 22 587–22 615, 2015.

[14] S. He, W. Lin, and S.-H. G. Chan, "Indoor localization and automatic fingerprint update with altered AP signals," *IEEE Transactions on Mobile Computing*, vol. 16, no. 7, pp. 1897–1910, Jul. 2017.

[15] Ó. Belmonte-Fernández, R. Montoliu, J. Torres-Sospedra, E. Sansano-Sansano, and D. Chia-Aguilar, "A radiosity-based method to avoid calibration for Indoor Positioning Systems A radiosity-based method to avoid calibration for Indoor Positioning Systems," *Expert Systems With Applications*, 2018.

[16] P. Krishnan, S. Krishnakumar, R. Seshadri, and V. Balasubramanian, "A robust approach for maintenance and refactoring of indoor radio maps," *Lecture Notes in Computer Science*, vol. 8487 LNCS, pp. 360–373, 2014.

[17] P. Wilk, J. Karciarz, and J. Swiatek, "Indoor radio map maintenance by automatic annotation of crowdsourced Wi-Fi fingerprints," *2015 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015*, Oct. 2015.

[18] S. H. Jung and D. Han, "Automated Construction and Maintenance of Wi-Fi Radio Maps for Crowdsourcing-Based Indoor Positioning Systems," *IEEE Access*, vol. 6, pp. 1764–1777, 2017.

[19] Y. Chen, W. Trappe, and R. P. Martin, "Attack detection in wireless localization," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*. IEEE, May 2007, pp. 1964–1972.

[20] C. Laoudias, M. P. Michaelides, and C. G. Panayiotou, "Fault detection and mitigation in WLAN RSS fingerprint-based positioning," *Journal of Location Based Services*, vol. 6, no. 2, pp. 101–116, Jun. 2012.

[21] A. Melkonyan, S. Yalamanchili, D. Akopian, and P. C. Chen, "Integrity monitoring and thresholding-based wlan indoor positioning algorithm for mobile devices," in *2011 6th International Conference on System of Systems Engineering*. IEEE, Jun. 2011.

[22] M. Maheshwari, S. Ananthanarayanan, A. Banerjee, N. Patwari, and S. K. Kasera, "Detecting malicious nodes in RSS-based localization," in *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*. IEEE, Jun. 2011.

[23] D. Zhang, G. Qiu, Y. Gao, X. Fang, R. Cheng, A. Chang, and C.-Y. Chan, "Crowdsourcing based radio map anomalous event detection system for calibration-on-demand," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Oct. 2014.

[24] Y.-C. Chen and J.-C. Juang, "Outlier-detection-based indoor localization system for wireless sensor networks," *International Journal of Navigation and Observation*, vol. 2012, pp. 1–11, 2012.

[25] J. Torres-Sospedra, P. Richter, G. M. Mendoza-Silva, E. S. Lohan, and J. Huerta, "Supporting Materials for Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates." [Online]. Available: https://doi.org/10.5281/zenodo.1305415

[26] D. Nüst, C. Granell, B. Hofer, M. Konkol, F. Ostermann, R. Sileryte, and V. Cerutti, "Reproducible research and GIScience: an Evaluation using AGILE conference papers," *PeerJ Preprints*, 2018.

[27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[28] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000*, vol. 2, 2000, pp. 775–784.

[29] A. G. Dempster, B. Li, and I. Quader, "Errors in determinstic wireless fingerprinting systems for localisation," in *2008 3rd International Symposium on Wireless Pervasive Computing*, May 2008, pp. 111–115.

[30] N. Marques, F. Meneses, and A. Moreira, "Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov. 2012, pp. 1–9.

[31] J. Torres-Sospedra and A. Moreira, "Analysis of sources of large positioning errors in deterministic fingerprinting," *Sensors*, vol. 17, no. 12, 2017.

[32] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-Term Wi-Fi fingerprinting dataset and supporting material," Nov. 2017. [Online]. Available: https://doi.org/10.5281/zenodo.1066041

[33] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "CRAWDAD dataset mannheim/compass (v. 2008-04-11)," Downloaded from https://crawdad.org/mannheim/compass/20080411, Apr. 2008.

[34] T. King, T. Haenselmann, and W. Effelsberg, "On-demand fingerprint selection for 802.11-based positioning systems," in *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Jun. 2008, pp. 1–8.

[35] S. Eisa, J. Peixoto, F. Meneses, and A. Moreira, "Removing useless APs and fingerprints from WiFi indoor positioning radio maps," in *International Conference on Indoor Positioning and Indoor Navigation*, Oct. 2013, pp. 1–7.

[36] E. Laitinen and E. S. Lohan, "Are all the access points necessary in WLAN-based indoor positioning?" in *2015 International Conference on Location and GNSS (ICL-GNSS)*, Jun. 2015, pp. 1–6.

[37] ——, "On the Choice of Access Point Selection Criterion and Other Position Estimation Characteristics for WLAN-Based Indoor Positioning," *Sensors*, vol. 16, no. 5, 2016.

[38] H. Al-Moukhles, A. K. Jaber, and I. Abdel-Qader, "Impact of APs selection scheme on compressive sensing-fingerprinting based IPS performance," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, Oct. 2016, pp. 1–7.

[39] M. Dashti and H. Claussen, "A metric to describe access point significance in location estimation," in *2016 13th Workshop on Positioning, Navigation and Communications (WPNC)*, Oct. 2016, pp. 1–6.

[40] T. Hayashi, D. Taniuchi, J. Korpela, and T. Maekawa, "Spatio-temporal adaptive indoor positioning using an ensemble approach," *Pervasive and Mobile Computing*, vol. 41, pp. 319–332, 2017.