Quantifying the Degradation of Radio Maps in Wi-Fi Fingerprinting

Ivo Silva*, Cristiano Pendão*, Joaquín Torres-Sospedra†, Adriano Moreira*
* Algoritmi Research Center, University of Minho - Portugal
†UBIK Geospatial Solutions S.L., Castellón, Spain

***Abstract*—One of the most common assumptions regarding indoor positioning systems based on Wi-Fi fingerprinting is that the Radio Map (RM) becomes outdated and has to be updated to maintain the positioning performance. It is known that propagation effects, the addition/removal of Access Points (APs), changes in the indoor layout, among others, cause RMs to become outdated. However, there is a lack of studies that show how the RM degrades over time. In this paper, we describe an empirical study, based on real-world experiments, to evaluate how and why RMs degrade over time. We conducted site surveys and deployed monitoring devices to analyse the radio environment of one building over 2+ years, which allowed us to identify significant changes/events that caused the degradation of RMs. To quantify the RM degradation, we use the positioning error and propose the RM degradation ratio, a metric to directly compare two RMs and measure how different they are. Obtained results show that the positioning performance is much better when RMs are collected on the same day as the test data, and although RM degradation tends to increase over time, it only leads to large positioning errors when significant changes occur in the Wi-Fi infrastructure, making previous RMs outdated.**

***Index Terms*—radio map, radio environment, Wi-Fi fingerprinting, degradation, radio signals, indoor propagation**

I. Introduction

Due to the proliferation of Wireless Local Area Networks (WLANs), Wi-Fi-based positioning systems have been used since the 2000s to locate and track users inside buildings. Wi-Fi fingerprinting, used in RADAR [1], is one of the most used techniques for indoor positioning based in WLAN. It is based on the idea that each indoor location is characterized by a unique set of Received Signal Strength Indicator (RSSI) values from existing Access Points (APs). The collection of several Fingerprints (FPs) associated with the Reference Points (RPs) where they were collected allows to build a Radio Map (RM). Then, a position estimate is obtained using an algorithm that compares an operational FP to the ones in the RM. Due to the high variability of Wi-Fi signals, Wi-Fi fingerprinting has an accuracy between $2\,\mathrm{m}$ to $8\,\mathrm{m}$ [2].

It is commonly assumed by the research community that one of the main drawbacks of Wi-Fi fingerprinting is the need to update the RM in order to maintain the performance of the system over time [3, 4]. The problem with the RM degradation is that it represents a snapshot of the radio environment at the time when the FPs were collected and, after some time, they no longer represent the reality, thus the RM becomes outdated and needs to be maintained. As stated in the literature [1, 5, 6, 7], some factors associated with RM degradation are: the propagation of radio signals indoors (reflection, scattering, multipath, etc); modifications in the indoor layout (moving furniture or other objects); the removal/displacement/addition of APs; and, the presence/absence of people and their Wi-Fi enabled devices, including Wi-Fi hotspots that often appear, change position, and disappear from the building.

RM maintenance can be accomplished by performing periodic manual site surveys, or by exploring methods for automatic maintenance, e.g., using Simultaneous Localization and Mapping (SLAM) [8] or collaborative approaches [9] where users contribute to the RM, among others [10, 11]. Despite those ways of updating the RM (manually or automatically), selecting the appropriate timing for the update is still an open issue. Detecting when a RM has actually degraded significantly –requiring to be updated in order to provide reliable indoor positioning– can be considered a major challenge.

By having several versions of the building's RM, it is possible to conduct an empirical analysis of the RM degradation over time. We analyse the Wi-Fi infrastructure to detect changes regarding the addition/removal of APs and observe how these events may affect the Wi-Fi fingerprinting performance using RMs collected. Then, we analyse the AP observation frequency, i.e., the number of times an AP is detected over the total number of scans, and determine why some APs are detected more or less often.

To conduct our study, we collected two types of long-term Wi-Fi fingerprinting datasets in the same building. The first is based on a comprehensive manual site survey and the second relies on several fixed Monitoring Devices (MDs), also known as anchors, that periodically collect Wi-Fi FPs. In addition, we used two metrics: the averaged positioning error and the Radio Map Degradation Ratio (RMDR). We propose the RMDR as a metric to measure the variations between two RMs. With both, we can assess whether the changes in the radio environment (observed in the WLAN analysis) lead to a poorer accuracy.

The main contributions of this paper are: 1) Analysis of the Wi-Fi infrastructure based on long-term datasets; 2) Radio map degradation ratio, a metric to quantitatively compare two RMs; 3) Real-world evaluation of RM degradation in a building, over 2+ years; 4) Study of the AP observation frequency, to assess which APs are more likely to be detected in Wi-Fi scans.

Corresponding Author: Ivo Silva (ivo@dsi.uminho.pt)
This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020 and the PhD fellowship PD/BD/137401/2018. J. Torres-Sospedra acknowledges funding from Torres Quevedo programme (PTQ2018-009981).

II. Related Work

The main focus of the research community regarding the degradation of RMs is on solutions to overcome the problem of the dynamically changing environment [3, 4, 11, 12]. Although these works assume that the RM tends to become outdated with time, they lack an analysis of the degradation of RMs. Instead, they try to propose methods to automatically build and maintain a RM.

In an effort to avoid repeating the site survey, which is laborious and time-consuming, some solutions explore fixed anchors collecting FPs to refresh the RM [10, 11, 13], while others [9, 14] explore crowdsourced data from users to perform automatic construction of RMs. SLAM has also been explored to perform the site survey [8, 15]. In addition, there are techniques to optimize the RM [3] hence, avoiding the RM maintenance process.

In [4], a brief analysis is presented about the signal strength variation of APs after 3, 6 and 9 days, and after 6 months. It shows a clear deterioration in the APs' signal strength after 6 months. This experiment showed that RMs tend to degrade over time but it did not justify how the RM degraded, whether the degradation was continuous, or if there was an event that caused such change.

In addition, this analysis only compared periodically collected data that goes up to 6 months, which can be in many cases a short period of time. The work in [7] proposes a method to detect RM degradation, by identifying changes in the APs' position in space. This is achieved by computing the average of the RSSI distance to the best match over multiple operational points. To detect RM degradation, alterations in the Wi-Fi infrastructure are simulated by manipulating the test data of the datasets. In [16], it is proposed an outlier detection system to identify RM degradation based on the detection of APs abnormal signal strength. It is capable of identifying events when an AP has suffered RSSI changes in time, however, it mostly focuses on short-term changes (over two days in the performed experiments). This excludes significant long-term changes that occur in the building.

Some tools, such as Netspot, TamoGraph Site Survey, Ekahau Connect[1], among others, allow to analyse the radio environment, e.g. to detect areas with a lower coverage or areas with interference. Being mostly focused on the network performance, these tools can be used to analyse the radio environment, but they do not provide ways of comparing the RM of the building over time nor comparing how they perform if they are old. Furthermore, these tools need manual surveying which is laborious to accomplish especially in large buildings.

Several Wi-Fi fingerprinting datasets have been made public aiming to improve research comparability and reproducibility [17, 18]. Most datasets available are collected during one day or a few consecutive days, thus both RM and test dataset are collected in close time proximity. To enable the evaluation of RMs as they become older, Mendoza-Silva [19, 20] proposed a long-term dataset. Periodic manual site surveys were performed every month to collect different types of datasets at a library building. This allowed to collect different RM versions over a period of 25 months, and evaluate the RMs over time. During the first month (June 2016), they collected 15 RMs, with 4 being collected on the same day.

[1] https://www.netspotapp.com, https://www.tamos.com/products/wifi-site-survey/, https://www.ekahau.com/solutions/wi-fi-heatmaps/

III. Approach

This paper builds on previous contributions but goes further. It considers two types of long-term datasets for evaluation of RM degradation, one based on manual site surveys to collect FPs in RPs, and another, based on MDs that continuously collect data in a period longer than 2 years. This allowed to perform an analysis of the Wi-Fi infrastructure to detect events that caused changes in the network and evaluate how they impacted the performance of the Indoor Positioning Systems (IPSs). RM degradation was quantified and evaluated in long term without manipulating the datasets to simulate alterations in the Wi-Fi infrastructure. Furthermore, an analysis of the AP's observation frequency was performed to model the behaviour of APs detected in FPs according to the RSSI.

We start this work following the same assumption as the research community, i.e., RMs tend to degrade over time, with the purpose of quantifying the RM degradation and finding whether the degradation of RMs is related to events that have caused significant changes in the radio environment. This section focuses on the definition of the RM, and the metrics for measuring RM degradation.

In this paper, the RM is defined as RM $= \{(\rho_1, ws_1), ..., (\rho_m, ws_m)\}$, which is the set of m FPs, each associated to the RP ρ where it was collected. Each FP, defined as $ws = \{RSSI_1, ..., RSSI_n\}$ represents the set of n RSSI values of the APs detected by the Wi-Fi interface.

A. Radio Map Degradation Ratio

The RMDR metric measures how much a RM has changed since it was initially collected by comparing it to a more recent one. It calculates the difference in estimated RSSI values at each position, assuming that two versions of the RM share the same RPs. Essentially, the RMDR defines the amount of RSSI variation (degradation) per RP per AP. In case there are multiple FPs per RP, they are averaged into one FP with averaged RSSI values. This allows to reduce noise from consecutive FPs. We define the RMDR between two RMs a and b (b older than a):

$$RMDR = \frac{\varpi_{a,b}}{N_{RP} \times N_{AP}} \qquad [\text{dBm}] \tag{1}$$

where N_{RP} represents the number of RPs, N_{AP} represents the number of APs detected in both RMs, and $\varpi_{a,b}$ represents the dissimilarity between pairs of FPs from RMs a and b collected at the same RP, defined as:

$$\varpi_{a,b} = \sum_{i=1}^{N_{RP}} \sum_{j=1}^{N_{AP}} \left| RSSI_{ij}^{a} - RSSI_{ij}^{b} \right| \tag{2}$$

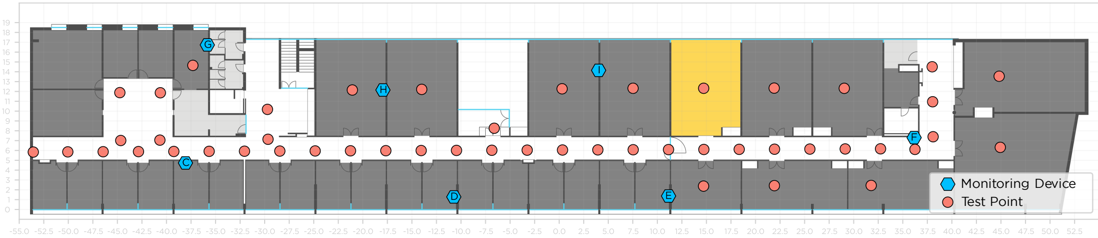

Fig. 1. DSI building: monitoring devices (blue hexagons) and test points (salmon circles).

where $RSSI_{ij}$ represents RSSI of AP_j in RP ρ_i for RMs a and b. In order to account for missing APs, whenever an AP is not detected in a FP its RSSI is assumed to be $-120\,\mathrm{dBm}$.

The RMDR is higher when the differences between RMs are also higher. Conversely, it is lower when both RMs are similar, hence it is zero when computed for the same RM, meaning that no degradation occurred.

B. Wi-Fi Fingerprinting

A complementary metric to evaluate the degradation of RMs is the positioning error (which is the ultimate goal). Wi-Fi fingerprinting has been used since the beginning of the century when RADAR was introduced [1]. For the estimation algorithm, we adopted the k-Nearest Neighbour (k-NN) method, where an operational FP is compared against the ones from the RM using a distance function. We opted to use the Manhattan distance, defined as:

$$d_M(ws^*, ws) = \sum_{i=1}^{N_{AP}} \left|RSSI^i_{ws^*} - RSSI^i_{ws}\right| \qquad (3)$$

where N_{AP} defines the number of detected APs, ws^* and ws represent the operational and RM FPs, respectively. If an AP is missing in either FP, a default RSSI of $-120\,\mathrm{dBm}$ is assumed. After computing the Manhattan distance between the operational and all the RM fingerprints, the estimated position is computed using k-NN, by obtaining the centroid of the k (with $k = 5$) most similar RM fingerprints:

$$\hat{\rho} = \frac{\sum_{i=1}^{k} \rho_i(x, y)}{k} \qquad (4)$$

where ρ_i represents the position of a ws.

IV. Long-term Wi-Fi Data Collection

In order to evaluate the degradation of RMs, we adopted two procedures to collect Wi-Fi FPs over time. The first is based on a site survey, where FPs were manually collected at the building. The second is based on MDs, also known as anchors, which were installed in known locations continuously collecting FPs. The manual site survey method has the advantage of containing Wi-Fi FPs from more reference locations, thus being a more dense RM. The method based on MDs has the advantage of being able to autonomously collect Wi-Fi samples without human intervention, over a long time.

A. Experiments Setup

Experiments were conducted at the Department of Information Systems (DSI) building located in the University of Minho's Azurém Campus. The building measures $106 \times 19\,\mathrm{m}$ and has several office rooms, laboratories, and classrooms. It is used daily by a couple of hundred people to attend classes and to conduct research activities. Fig. 1 depicts the floor plan of the building, the test points (RPs where FPs were collected in the site survey), and the MDs (used for continuous collection of FPs). The DSI building is equipped with multiple Wi-Fi APs to ensure that all users of the building can access the internet efficiently. These APs (Cisco Aironet Series) are easily identified in Wi-Fi scans due to their SSID being the same in the entire campus, however many other networks are also found in Wi-Fi scans. These networks are usually associated with printers, mobile APs, or other APs installed in laboratories and office rooms.

B. Manual Site Survey

On a period over 2 years, 12 datasets were obtained using a Raspberry Pi 3B+. Site surveys in which datasets were obtained are represented in blue in Fig. 5. Each dataset is composed of 20 FPs at each testing point (see Fig. 1) and can be seen as a different version of the building's RM. Unfortunately, due to the pandemic, it was not feasible to perform the manual site surveys during most of 2020.

Although it is a time-consuming and laborious task, the datasets resulting from this work can be directly compared to evaluate the RM degradation. Since these datasets contain samples from a wider set of test points (a total of 49), they better represent the building's radio environment than the datasets obtained with MDs (only 7 in total). In order to enrich these datasets, in addition to the manually surveyed points, we also considered FPs from MDs collected around the same time that these datasets were obtained. In summary, 12 datasets were collected, each integrating a total of 13 440 FPs collected across 56 distinct positions.

C. Monitoring Devices Data

Although the installation of infrastructure to collect data tends to be expensive, we chose the Raspberry Pi 3B+ to implement the MD. The main advantages of this device are that it is low-cost, it already includes 2.4 and 5 GHz Wi-Fi interfaces (802.11b/g/n/ac), it is simple to configure

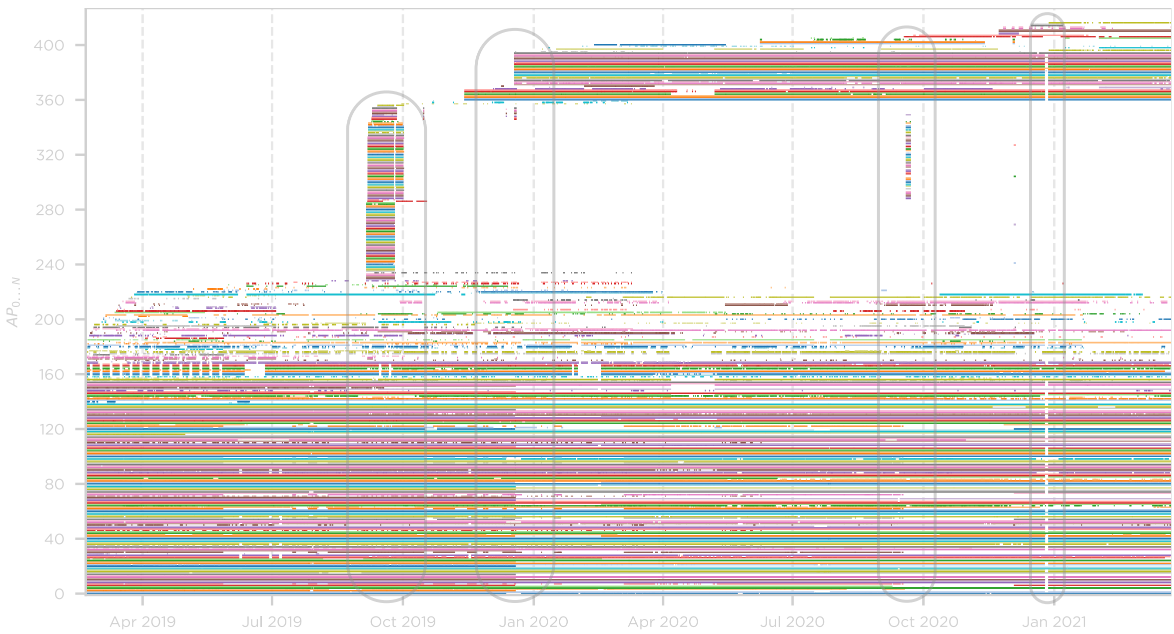

Fig. 2. APs detected by the MDs over time.

and deploy, and it is based on Linux which facilitates the development of software modules. We devised a Java program to collect a new FP every 60 seconds. In order to avoid getting outdated RSSI values, the program performs two scans and considers the last one, ensuring that the previously buffered FP is replaced by new RSSI values. The program sends FPs to a centralized server that gathers data from all MDs and also saves the data locally into an SQLite database.

After testing the data collection program, seven MDs were deployed in the DSI building (Fig. 1) on 2019-02-19 and started collecting data on that day. Since the MDs are continuously collecting Wi-Fi FPs, one can easily obtain the RM of the building based on MD data, at any time period. Although it has a low number of RPs (only 7) for the size of the building considered, this RM can be used with Wi-Fi fingerprinting as a metric to evaluate if the positioning performance degrades when the RM is older. MD data was used to create RMs in two ways: to complement the manual site surveys, and to generate RMs integrating MD data only. The blue and yellow marks in Fig. 5 represent the dates when these data were obtained. More than 7 million FPs were collected from all MDs since they started collecting data.

V. Radio Environment Over Time

The continuous monitoring provided by MDs allows to observe variations in the radio environment over time. Fig. 2 shows all visible APs, since we started this work, in 2019. This plot aggregates data from all MDs deployed in the building, hence, it shows all visible APs over the analysed period. APs are sorted according to the first time of detection. Since some MDs are deployed near windows, they can detect APs from neighbour buildings, which are also included in Fig. 2. Each line, represented in a different colour, depicts a period during which an AP was detected. A total of 417 unique APs were detected during this period, considering that each unique Media Access Control (MAC) address counts as one AP. A total of 4 448 APs that were rarely observed (in less than 1% of FPs) were ignored and are not displayed in the plot.

Fig. 2 serves as a tool for analysing changes in the radio environment. It shows that there are some APs that were detected over the entire observation period. Some APs have intermittent behaviour (around AP_{160}), being switched on and off several times. There is also the case where APs only show during a short period of time (in Oct. 2019 and Sept. 2020). Some APs appear at a specific time (Dec. 2019) and remain visible until the end of the data collection. The variations observed in Fig. 2, demonstrate that the radio environment is highly dynamic, including many short-term variations and several significant changes in the Wi-Fi infrastructure over time. A description of the more significant events (highlighted in grey in Fig. 2) is presented next.

1) Sept. 2019: A set of new APs are detected over one month, then disappear and are not observed again. Upon a deeper analysis through the FPs obtained during that period, we observed that many new APs, associated with the University's WLAN infrastructure, were detected during this period.

2) Dec. 2019: Many APs stopped being detected and several new ones started being detected around the same time. Due to the significant amount of APs that changed in this period, this is a dramatic change in the radio environment.

3) Sept. 2020: We see similar behaviour to the one observed in Sept. 2019, but the newly added APs were detected during a shorter period.

4) Dec. 2020: Immediately before Jan. 2021, almost all APs stopped being detected during a couple of days (vertical white space around the time indicating that the vast majority of APs was switched off). This suggests that there was a power outage that caused APs to be down during this period. APs return to normal operation after this period. See Section VII for further details.

5) Other events: Many other variations and periods in which APs are switched on and off can be found in Fig. 2, demonstrating that the radio environment is highly dynamic not only in the long-term but also in short-term. These short-term variations are probably associated with the building's activity where people use their APs and hotspots, which are easily moved, switched on and off, hence they are not as constant as the APs from the university's WLAN infrastructure.

VI. Radio Map Degradation Over Time

Whether or not the highly dynamic radio environment has an impact on IPSs is analyzed in this section, where the results regarding the RM degradation over time, using the mean positioning error and the RMDR metrics, are presented.

A. Manual Site Surveys Data

Each dataset was used as RM and all subsequent datasets (the ones obtained after the RM) were used as test datasets. This allows the evaluation of the RM over time since the day it was initially collected. For example, the 2019-12-11, 2020-01-15, 2020-02-19, 2021-04-23 datasets were used as test data when considering the 2019-12-11 dataset as RM. In the case where the RM and test datasets are the same (when the RM is tested on the same day it was created), we have selected 5 FPs as test data and the remaining 15 as RM data, for each testing point. During the majority of 2020, it was not possible to perform the manual site survey due to COVID.

Fig. 3 (a) depicts the mean error achieved for each version of the RM. Each line in the plot represents the times when a RM was evaluated, over time. As expected, better results are obtained when the RM is more recent. When the RM is obtained on the same day as the test dataset, the results are very good. The plot shows that the degradation is not incremental over time and it is not guaranteed that time itself is the reason for an increase in the RM degradation. In general, we can see that older RMs have similar performance as they get older, for instance, between Sept. 2019 and Dec. 2019.

The positioning results obtained with RMs and testing data collected on the same day are much better than results obtained in more realistic conditions where testing data is collected days or months after the RM was created. Many results reported in the literature might suffer from this problem, thus not representing the real performance of the IPSs.

Furthermore, results show that Sept. 2019 was not a good period to build the RM because, as described in Section V, during this period many new APs were added and removed about a month later. All tests with the Sept. 2019 RM had worse performance than older RMs. This shows that when the RM contains a set of APs that are removed, it affects the positioning results based on Wi-Fi fingerprinting since the implemented version assigns a default RSSI to APs that are not detected.

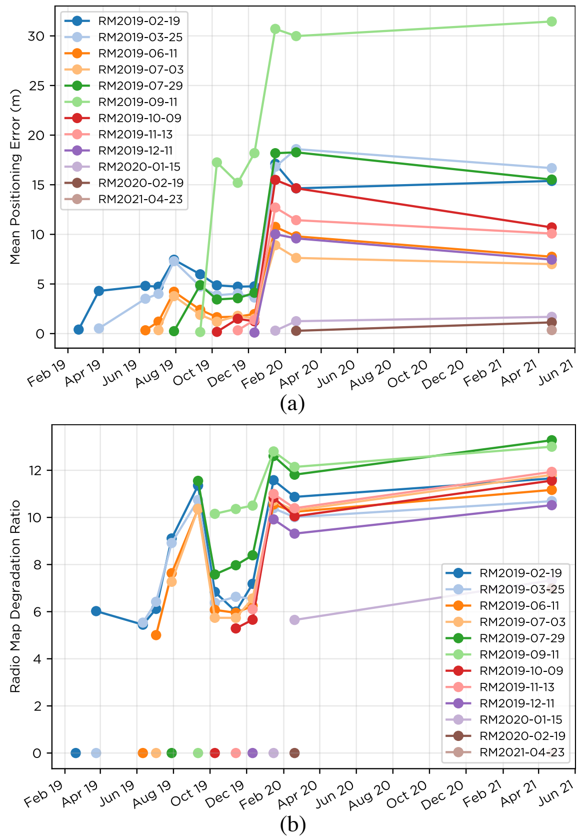

Fig. 3. Manual site survey metrics: (a) mean error of Wi-Fi fingerprinting; (b) RMDR.

A clear degradation of results in all RMs collected before Jan. 2020, is observed in 2020. As shown in Fig. 2, in Dec. 2019, there were many APs that were disconnected, and new ones were added, causing significant changes in the radio environment. Therefore, all RMs that were created before Jan. 2020 became outdated, as shown by the worse positioning results in 2020.

Despite a very long time between the last two datasets (2020-02-19 and 2021-04-23), the RMs that were obtained in 2020 did not have significantly worse results with the 2021 dataset. This is justified by the reduced activity in the building and almost no changes in the Wi-Fi infrastructure during 2020.

The RMDR was evaluated for each RM version by comparing it with itself and the subsequent ones. Fig. 3 (b) shows the RMDR of each RM over time. When compared with itself, the RMDR is zero, as shown in the plot. The RMDR achieved for all RMs follows a similar behaviour as the one observed in Fig. 3 (a), except for the values observed in 2019-07-29 and 2019-09-11. During this period, the RMDR increased but it did not translate into significantly worse positioning results. The RMDR rose substantially in the RMs after Dec. 2019, following the same behaviour observed in Fig. 3 (a). This shows that changes that occurred in the radio environment after Dec. 2019 are different than the ones observed in Jul. and Sept. 2019 (where the RMDR was higher, but the mean error remained almost the same). Our analysis of the building's Wi-Fi infrastructure (see Section VII) shows that these changes are related to distinct events.

In sum, these results demonstrate that better results are achieved when the RM and test dataset are collected on the same day. When a RM is tested after some time, there is a degradation in the positioning performance. RM degradation is directly related to changes in the Wi-Fi infrastructure, especially when several APs are removed from the building.

B. Monitoring Devices Data

We adopted the same approach as the one in Section VI-A in order to evaluate RM degradation using Wi-Fi fingerprinting and the RMDR metric. As a result of being continuously collecting Wi-Fi FPs, it allowed to obtain more datasets in comparison to the manual site surveys. In this experiment, we considered MDs datasets on the same dates as manual site surveys experiments, to validate whether the results are similar. MDs datasets include data from months during which the university was closed or partially opened, due to COVID.

Fig. 4 (a) depicts the Wi-Fi fingerprinting mean error over time, for each RM version. These results are in accordance with the ones of Fig. 3 (a). Again, the Sept. 2019 RM is only valid during the period when it was collected. It has a large positioning error in all other tests. The event from Dec. 2019 significantly affected the results of RMs that were built before that. In 2020, we see that new RMs created after Dec. 2019 have good performance results, demonstrating that no significant changes in the radio environment have occurred during the year. In this period, the mean error is low because only 7 test points (MDs positions) were considered, and the building's activity was substantially reduced due to COVID restrictions. Regarding the RMDR, Fig. 4 (b) shows that RMs from MDs also register high RMDR values in Sept. 2019 and Jan. 2020, having a similar behaviour to the results achieved with the site survey RMs (Fig. 3 (b)).

In Sept. 2019, a significant change in the radio environment has occurred, represented by the peak in the RMDR for all RMs. The RMDR is lower for the RMs created after Dec. 2019, showing that the radio environment has changed significantly during that period.

VII. Events Timeline

In order to establish a cause/effect relationship between the changes in the radio environment (identified in Section V) and the obtained results (in Section VI), we tried to identify the major events that occurred during the data collection period (see Fig. 5). This timeline includes times when there were interventions in the building's Wi-Fi infrastructure, periods during which the university was closed due to COVID, and times (blue and yellow marks) when datasets were obtained. Blue and yellow marks represent times when both types of datasets were collected. Yellow marks represent dates when

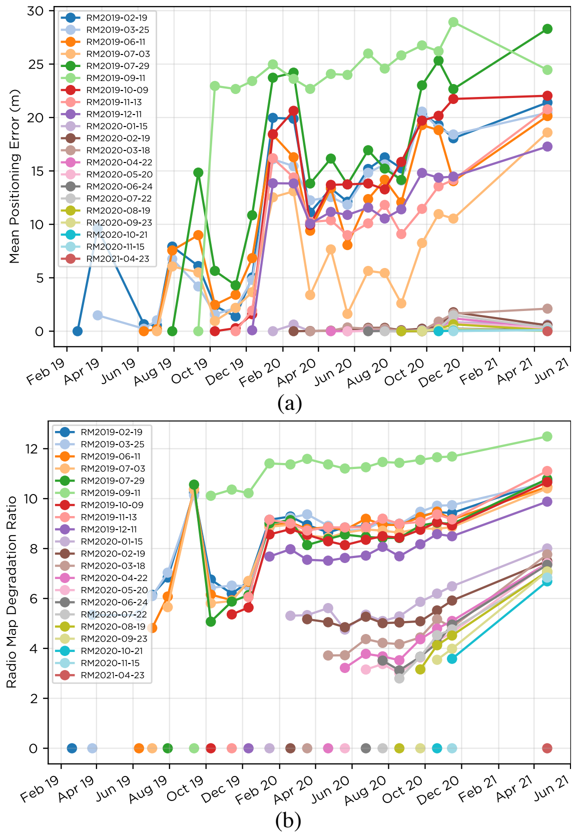

Fig. 4. Monitoring devices metrics: (a) mean error of Wi-Fi fingerprinting; (b) RMDR.

only MD datasets were collected. The last time a manual site survey was conducted was on 2021-04-23, a few days after the university re-opened after being closed due to COVID restrictions. No datasets from MDs were obtained between Nov. 2020 and Apr. 2021, because one of the devices stopped working, and was only fixed in Apr. 2021. The main interventions in the building's Wi-Fi infrastructure, which affected the results (see Section VI), are described next:

- In Sept. 2019, with the beginning of the school year, the technicians performed several tests in the network, during which they used the existing APs to transmit several new SSIDs which were not previously detected. A similar test was also conducted in Sept. 2020[2].
- The building's WLAN suffered a maintenance intervention in Dec. 2019. New APs with a larger user and traffic capacity were installed in classrooms (right side in Fig. 1), and some of the replaced APs were moved into the office areas to increase coverage (left side in Fig. 1).
- In Dec. 2020, a power outage caused problems in the building's power supply and a significant number of APs were switched off for a couple of days. This caused the network to be down, and since it occurred around

[2]Information about operations in the Wi-Fi infrastructure was obtained from the University services.

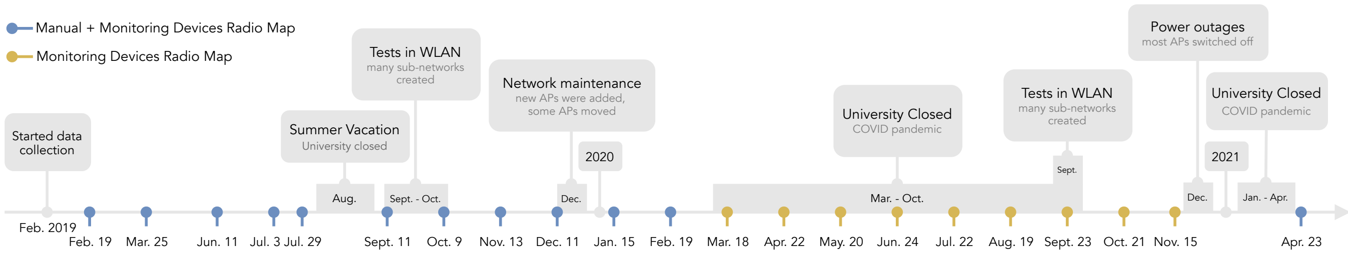

Fig. 5. Significant events in DSI building over time.

Christmas time, the technicians were only able to solve this problem after a couple of days. Although the building suffered power outages, the circuit that powers MDs seems to be independent of the one that powers the APs, since we can see that there are a few APs that are still visible during this period (Fig. 2), hence MDs were still operating and some APs were also working. We suppose that the APs detected during this period are either rogue APs (not from the university's dedicated WLAN), or APs from neighbour buildings, which can be detected by MDs. After this issue was solved, the APs returned to their previous state and were detected by MDs.

VIII. APs Observation Frequency

The FPs collected by MDs allowed us to conduct an analysis of the APs observation frequency. APs are not always detected in Wi-Fi FPs due to being far from the receiver device, or due to characteristics of signal propagation. When an AP is missing in a FP, it is either ignored or assigned a default value. Usually, this value serves as a penalization in the localization algorithm for the AP not detected in the FP.

Researchers tried to characterise the APs RSSI distribution [11, 21] by showing histograms of AP's signal levels over a time window. Although there are some conflicting conclusions about which model better fits RSSI distribution, most authors agree that RSSI histograms resemble a Gaussian distribution in most cases. Many other distributions are also considered in the literature due to the diversity of RSSI histograms. Despite performing an analysis of the signal distribution in RSS space, these works miss an important aspect which is the ratio of the times the AP is detected over the total number of FPs analysed. This aspect is important especially in solutions that perform AP selection, either to improve positioning performance or to reduce computational effort [22].

In order to characterize APs according to the ratio of times they are detected in Wi-Fi FPs, we defined the AP observation frequency as:

$$r = \frac{n}{N} \tag{5}$$

where n represents the number of times the AP was detected over the total of N FPs considered.

To analyse AP observation frequency, we considered FPs collected from all MDs during one day. Each MD detects unique APs, and for each MD, one can obtain the observation frequency of each detected AP in the considered set of FPs. In order to understand whether the APs observation frequency is affected by signal strength, we also obtained the mean and Standard Deviation (stdev) for each AP.

Each AP detected by MDs in the considered period is represented by a point in Fig. 6. Each AP is represented according to its observation frequency, mean RSSI, and RSSI stdev defined by the colour of each point. Fig. 6 shows that there is a moderate positive correlation between the mean RSSI and the observation frequency (Pearson correlation $r = 0.566$), demonstrating that APs with higher RSSI values are detected more often. This is expected because for APs whose signal is weak are detected fewer times due to the receiver sensitivity and variations in signal propagation. In this particular case, the receiver sensitivity of the Raspberry Pi 3B+ is $\approx -92\,\mathrm{dBm}$, since there is no AP detected whose RSSI is below that value. Since the Pearson correlation $p < 10^{-4}$, it rejects the null hypothesis, therefore these results have statistical relevance.

Regarding the RSSI stdev of APs, we see a higher concentration of APs whose RSSI stdev is close to zero when the mean RSSI is low. This is probably related to the number of FPs in which an AP is detected, being that, when the AP has lower signal strength, it is detected fewer times (low observation frequency), which consequently leads to lower RSSI stdev. There is no clear tendency regarding the APs whose RSSI stdev is high, because, these APs appear in areas with low and high observation frequency.

We observed similar results using the same approach in other periods, a few months apart. Also, we conducted the same experiment considering all FPs collected by MDs over a week, and the obtained results are in line with the ones presented in this document. Those results are omitted in this paper due to space constraints.

This study is important to support works that model the behaviour of APs. Many models, such as the Log-Distance Path Loss (LDPL) model ignore the probability of an AP not being detected in a FP. This empirical study shows that, in order to accurately model the behaviour of radio signals at the receiver, propagation models should account for the AP's observation frequency according to the signal strength, being that APs whose signal is stronger are detected more often.

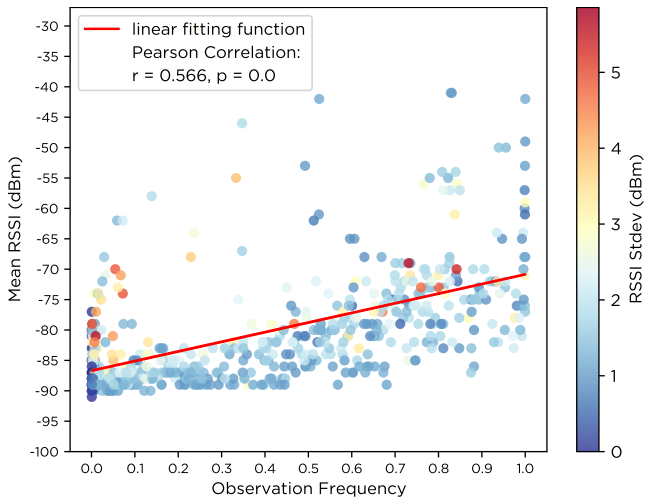

Fig. 6. APs observation frequency vs. RSSI mean and stdev.

IX. Conclusions and Future Work

In this paper, we conducted an empirical study on the degradation of RMs at an office/classroom building. We collected long-term Wi-Fi data, by performing several manual site surveys and deploying multiple MDs that autonomously collected Wi-Fi data.

The long-term RM degradation was assessed with the positioning error and the proposed RMDR. Achieved results showed that, although there is a tendency to increase over time, the RM degradation is mostly caused by dramatic changes in the Wi-Fi infrastructure. The positioning performance is significantly affected when many APs are removed from the building, making all RMs created before this change, outdated. Therefore, the RM should be updated after significant changes occur in the Wi-Fi infrastructure. The RMDR measures how different two RMs are by comparing existing APs and their signal levels for each RP. The RMDR can be used to detect RM degradation, but it is not an indicator of worse positioning performance. When just a few APs are added to the Wi-Fi infrastructure, it does not affect the positioning performance, but still leads to an increase of the RM degradation defined by the RMDR. Also, the positioning results obtained with training and testing data collected on the same day might be misleading by providing unrealistically good performance.

Regarding the APs observation frequency, we found that there is a positive correlation between signal strength and AP observation frequency, i.e., APs whose signal is stronger are detected more often than APs whose signal is weaker.

Plans for our future work include: explore other functions for the RMDR; evaluate this method in other scenarios; development of a solution to automatically detect RM degradation; continue with the MDs data collection; publish the datasets analysed in this paper; explore other fitting functions for the APs observation frequency; and, analyse the variation of day-to-day positioning performance to evaluate how short-term variations in the radio environment affect the results.

References

[1] P. Bahl and V. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system,” in *Proc. IEEE INFOCOM 2000. Conf. Comput. Commun. Ninet. Annu. Jt. Conf. IEEE Comput. Commun. Soc. (Cat. No.00CH37064)*, vol. 2, 2000, pp. 775–784. arXiv: 1106.0222.

[2] S. He and S.-H. G. Chan, “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,” *IEEE Commun. Surv. Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[3] R. Montoliu, E. Sansano, O. Belmonte, *et al.*, “A New Methodology for Long-Term Maintenance of WiFi Fingerprinting Radio Maps,” in *2018 Int. Conf. Indoor Position. Indoor Navig.*, IEEE, Sep. 2018, pp. 1–7.

[4] C. Wu, Z. Yang, and C. Xiao, “Automatic Radio Map Adaptation for Indoor Localization Using Smartphones,” *IEEE Trans. Mob. Comput.*, vol. 17, no. 3, pp. 517–528, Mar. 2018.

[5] J. S. Seybold, *Introduction to RF Propagation*, 1-3. Hoboken, NJ, USA: John Wiley & Sons, Inc., Sep. 2005, vol. 14, pp. 145–173.

[6] K. Kaemarungsi and P. Krishnamurthy, “Properties of indoor received signal strength for WLAN location fingerprinting,” *Proc. MOBIQUITOUS 2004 - 1st Annu. Int. Conf. Mob. Ubiquitous Syst. Netw. Serv.*, pp. 14–23, 2004.

[7] J. Torres-Sospedra, P. Richter, G. Mendoza-Silva, *et al.*, “Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates,” *IPIN 2018 - 9th Int. Conf. Indoor Position. Indoor Navig.*, no. September, pp. 24–27, 2018.

[8] R. Liu, S. H. Marakkalage, M. Padmal, *et al.*, “Collaborative SLAM Based on WiFi Fingerprint Similarity and Motion Information,” *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1826–1840, 2020. arXiv: 2001.02759.

[9] Y. Luo, Y. P. Chen, and O. Hoeber, “Wi-Fi-Based Indoor Positioning Using Human-Centric Collaborative Feedback,” in *2011 IEEE Int. Conf. Commun.*, IEEE, Jun. 2011, pp. 1–6.

[10] C. Pendao and A. Moreira, “FastGraph Enhanced: High Accuracy Automatic Indoor Navigation and Mapping,” *IEEE Trans. Mob. Comput.*, vol. 20, no. 3, pp. 1027–1045, Mar. 2021.

[11] Jie Yin, Qiang Yang, and L. Ni, “Learning Adaptive Temporal Radio Maps for Signal-Strength-Based Location Estimation,” *IEEE Trans. Mob. Comput.*, vol. 7, no. 7, pp. 869–883, Jul. 2008.

[12] H. Wang, L. Ma, Y. Xu, *et al.*, “Dynamic radio map construction for WLAN indoor location,” *Proc. - 2011 3rd Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2011*, vol. 2, pp. 162–165, 2011.

[13] L. Ni and A. Patil, “LANDMARC: indoor location sensing using active RFID,” *Proc. First IEEE Int. Conf. Pervasive Comput. Commun.*, pp. 407–415, 2003.

[14] D. Matos, A. Moreira, and F. Meneses, “Wi-Fi fingerprint similarity in collaborative radio maps for indoor positioning,” in *Inforum 2014*, 2014, pp. 184–194.

[15] P. Mirowski, R. Palaniappan, and T. K. Ho, “Depth camera SLAM on a low-cost WiFi mapping robot,” *2012 IEEE Conf. Technol. Pract. Robot Appl. TePRA 2012*, vol. 1, no. 908, pp. 1–6, 2012.

[16] D. Zhang, G. Qiu, Y. Gao, *et al.*, “Crowdsourcing based radio map anomalous event detection system for calibration-on-demand,” in *2014 Int. Conf. Indoor Position. Indoor Navig.*, IEEE, Oct. 2014, pp. 220–229.

[17] A. Moreira, J. Torres-sospedra, I. Silva, *et al.*, *Wi-Fi Fingerprinting dataset with multiple simultaneous interfaces*, 2019.

[18] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, *et al.*, “Wi-Fi crowdsourced fingerprinting dataset for indoor positioning,” *Data*, vol. 2, no. 4, pp. 1–16, 2017.

[19] G. Mendoza-Silva, P. Richter, J. Torres-Sospedra, *et al.*, “Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning,” *Data*, vol. 3, no. 1, p. 3, Jan. 2018.

[20] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, *et al.*, *Long-Term Wi-Fi fingerprinting dataset and supporting material*, Zenodo, Apr. 2020.

[21] J. Luo and X. Zhan, “Characterization of Smart Phone Received Signal Strength Indication for WLAN Indoor Positioning Accuracy Improvement,” *J. Networks*, vol. 9, no. 3, pp. 739–746, Mar. 2014.

[22] S. Eisa, J. Peixoto, F. Meneses, *et al.*, “Removing useless APs and fingerprints from WiFi indoor positioning radio maps,” in *Int. Conf. Indoor Position. Indoor Navig.*, IEEE, Oct. 2013, pp. 1–7.