A Microscopic Look at WiFi Fingerprinting for Indoor Mobile Phone Localization in Diverse Environments

Arsham Farshad, Jiwei Li, Mahesh K. Marina
The University of Edinburgh

Francisco J. Garcia
Agilent Technologies

***Abstract*—WiFi fingerprinting has received much attention for indoor mobile phone localization. In this study, we examine the impact of various aspects underlying a WiFi fingerprinting system. Specifically, we investigate different definitions for fingerprinting and location estimation algorithms across different indoor environments ranging from a multi-storey office building to shopping centers of different sizes. Our results show that the fingerprint definition is as important as the choice of location estimation algorithm and there is no single combination of these two that works across all environments or even all floors of a given environment. We then consider the effect of WiFi frequency bands (e.g., 2.4GHz and 5GHz) and the presence of virtual access points (VAPs) on location accuracy with WiFi fingerprinting. Our results demonstrate that 5GHz signals are less prone to variation and thus yield more accurate location estimation. We also find that the presence of VAPs improves location estimation accuracy.**

I. Introduction

In this paper, we take a microscopic look at the well-known WiFi fingerprinting approach when applied for indoor mobile phone localization. Specifically, we examine the impact of various aspects underlying a WiFi fingerprinting system, including: the definition of a fingerprint, run-time location estimation algorithms, frequency band and presence of virtual access points (VAPs). Our investigation considers several different real indoor environments ranging from a multi-storey office building to shopping centers of different sizes. Seven different definitions of fingerprints are considered that span RSSI based, AP visibility based and combinations of both. With respect to location estimation algorithms, we compare three different deterministic techniques (including the often used Euclidean distance based nearest neighbor method) with two probabilistic techniques that use Gaussian and Log-normal distributions for RSSI modeling.

Our findings are summarized as follows:

- Section IV: Our analysis shows that the fingerprint definition is at least as important as the choice of location estimation algorithm; the latter has received significantly more attention in the literature till date. Moreover, there is no single combination of fingerprint definition and localization algorithm that always yields the optimum localization result across all the different environments we considered. In fact, even different floors within the same building have different optimum combinations.
- Section V: We consider the impact of frequency band used (2.4GHz vs. 5GHz) on WiFi fingerprinting and find that 5GHz offers relatively better location accuracy due to lower RSSI variation.
- Section VI: We also consider, for the first time, the effect of virtual access points (VAPs), which are now becoming commonplace in most indoor environments. Contrary to intuition, we find that the presence of VAPs significantly improves WiFi fingerprinting accuracy which we believe is due to two reasons: VAPs have a substantial influence on the AP density, a factor known to affect accuracy with WiFi fingerprinting; and fingerprints obtained from different co-located VAPs operating on the same channel are somewhat dissimilar, capturing the temporal variability inherent to wireless signal propagation and providing robustness against it.

II. Related Work

WiFi fingerprinting has emerged as a popular WiFi based localization technique in the past 10-15 years since the idea was first put forth in the RADAR system [1]. The attractive thing about WiFi based localization approach is that it exploits the prevalent WiFi infrastructure in many indoor environments and the presence of WiFi interfaces now common in smartphones. With fingerprinting there is the added advantage of not having to go through the process of accurate radio propagation modelling which can be quite challenging in multipath rich indoor environments. Instead the idea is to use the signal characteristics at each location (usually signal strength from visible APs) as a signature to infer location. Generally speaking, fingerprinting systems consist of two phases. The first phase involves building a fingerprint database or constructing a radio map through measurements associated with known locations. This phase is sometimes referred to as site survey / offline / training phase. Then in the second phase, variously referred to as online / runtime / positioning / tracking phase, signal measurement samples collected by a user's device are used to "look up" the closest matching samples in the database / radio map to infer the user's location. Early WiFi fingerprinting systems including RADAR [1] and Horus [2] rely on an initial training phase to construct fingerprint database for use as a

This work was supported in part by a Cisco Research Award.

978-1-4673-1954-6/12/$31.00 © 2012

reference in the positioning phase later but training phase can be quite time consuming and expensive. More recent WiFi fingerprinting systems make this training phase automated via crowdsourcing using various mechanisms with increasing sophistication (e.g., Redpin [3], OIL [4], Zee [5]).

More closely relevant to this paper are the studies comparing different WiFi fingerprinting techniques (e.g., [6]) and analyzing the properties of WiFi signals as they pertain to location fingerprinting (see [7] and references therein). A number of factors are now recognized to have an impact on the accuracy of WiFi fingerprinting systems to varying degrees, including user orientation, temporal and spatial variations of WiFi signals, device hardware, transmit power, number of measurement samples [1], [2], [7].

Our work differs from and advances the previous work in the sense that it considers factors such as fingerprint definition, effect of frequency band and VAPs that are beyond those have been previously considered in the context of smartphone based WiFi fingerprinting in diverse environments. Concerning our investigation on the effect of frequency band, [8] have also come up the same conclusions although they do not analyze the underlying reasons. Specifically, [8] studied the effect of different device hardware types on RSSI behavior including some dual-band WiFi interfaces. The authors observed that 5GHz exhibits relatively low standard deviation of RSSI and they conjecture that it could be possibly be due to low interference and propagation effects in 5GHz band without any experimental validation.

III. Methodology

A. Data Collection

We obtain WiFi fingerprinting data for our study using Android phones and IndoorScanner, a custom mobile application we developed for this specific purpose. For each measurement position, which we note as the ground truth, IndoorScanner relies on the Android API (specifically, the getScanResults() method in the WifiManager class) to do multiple (20) scans, each taking approximately 1 second. Information gathered from each scan includes service set identification (SSID), basic service set identification (BSSID), RSSI, channel and UNIX timestamp. Scan results are annotated with the corresponding ground truth position and stored in a MySQL database, in a separate table for each different environment. We use either Samsung Galaxy S3 or HTC Nexus One phones, both Android based, to generate the various datasets.

B. Environments

We consider a multi-storey office building and three different shopping centers as representative set of diverse environments. Layout of these different environments is shown for reference in Figure 1 and Figure 2.

Multi-storey office building. As a representative office building, we consider the Informatics Forum building in the University of Edinburgh which houses the School of Informatics. We focus on five floors of this building which constitute the main areas with staff/student offices, common spaces and labs. Figure 1 shows the floor plan for two of the floors. Note that the grey area in the middle is empty across all floors. Also note that two of the floors, including the second floor shown in Figure 1(b), are slightly different with an open plan common space in place of some rooms. As a result the number of sampled measurement locations are different between floors — floors with open spaces have more number of measurement locations. There is a university run wireless LAN service across the whole building with several APs installed per floor. Each of these physical APs function as two virtual APs corresponding to two wireless networks with different user authentication mechanisms. In addition, a number of other APs can be seen across the building, some installed by various research groups in the building while others from surrounding buildings. The WiFi fingerprint dataset for this building was generated by measurements using our IndoorScanner app described above along the corridors and in common spaces at a granularity of 1 square meter cells, colored cyan in Figure 1.

Shopping centers. Besides the office building described above, we also consider three shopping centers of different sizes in Edinburgh, UK as shown in Figure 2. We use WiFi scan results with our IndoorScanner app along with a distinct id we manually assigned for each measurement position (shown as purple colored cells in Figure 2 to produce the individual datasets for each of these environments. Note that compared to the office environment described above, sampling of these shopping environments is sparser as they are public spaces with less flexibility in choosing measurement location and also given their size. These measurements were collected during busy shopping times to better capture a realistic usage scenario.

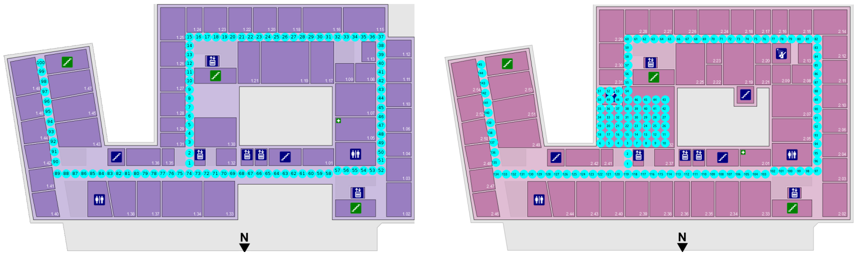
C. Fingerprint Definitions

What constitutes a WiFi fingerprint, i.e., the fingerprint definition, potentially influences the accuracy of a WiFi fingerprinting system even if other aspects such as the location estimation algorithm are kept fixed.

As a starter, a vector of mean[1] signal strength values from different WiFi APs seen at a location can be taken as the WiFi fingerprint for that location, as in [1]. We refer to this fingerprint definition as the *Default* fingerprint definition in the rest of this paper. However, as shown in section IV, we observe that this default definition yields poor location accuracy when compared to some of the alternative and "shorter" fingerprint definitions we consider in our study (7 in total). These other definitions are outlined below and share a common characteristic that they involve choosing a subset of APs (5 in our implementation) for each location that satisfy a particular criterion (e.g., highest strength).

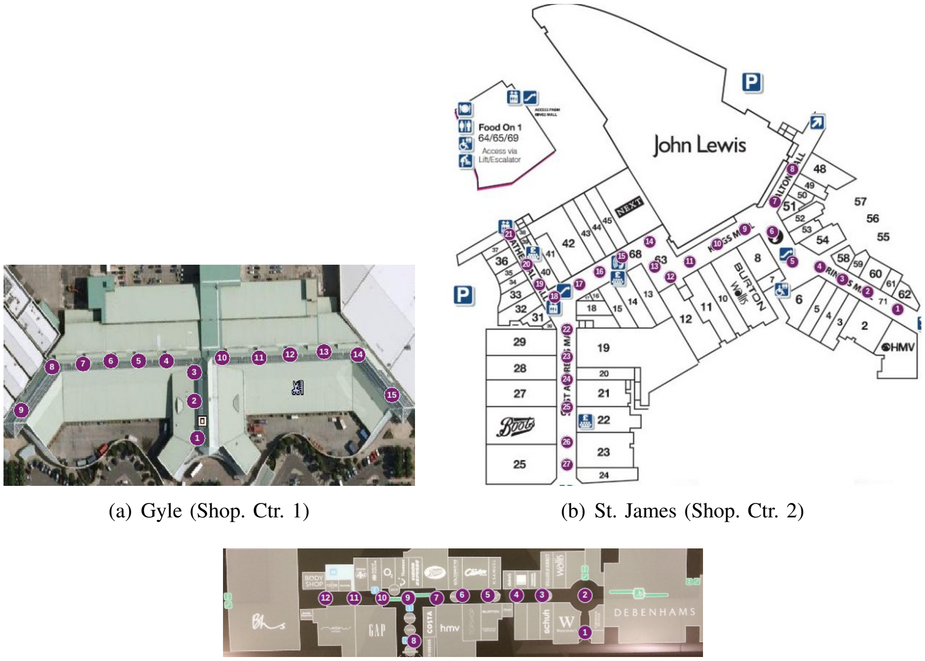
1) RSSI based: Received signal strength (RSSI) of beacons from APs is a key feature commonly considered in WiFi fingerprinting. We consider the following three different fingerprint definitions based on RSSI:

[1] This could be some other summary statistic (e.g., median).



(a) First Floor (b) Second Floor

Fig. 1: Floor plans for first and second floors of Informatics Forum, University of Edinburgh (office environment). Sampled locations during data collection are shown as cyan colored cells.



(a) Gyle (Shop. Ctr. 1) (b) St. James (Shop. Ctr. 2)

(c) Ocean Terminal (Shop. Ctr. 3)

Fig. 2: Layouts of three shopping centers in Edinburgh (shopping center environments). Purple colored cells represent the locations sampled during data collection.

Strength: In this definition, for each location in the training data, the subset of APs with *highest* mean RSSI values are considered along with the runtime fingerprint data corresponding to those chosen APs for estimating the location using one of the algorithms described in section III.D.

Stability: This definition focuses on the most stable subset of APs based on the standard deviation of their RSSI values. The rationale for considering this definition is two-fold: (i) as received signal strength is inherently time varying, the signals that vary less would more likely result in better accuracy of localization; (ii) [7] conclude from their analysis that RSSI standard deviation is the most influential factor determining the accuracy of a WiFi fingerprinting system.

Variance: This definition is based on the observation that it is ideal for a fingerprinting based localization system if fingerprints from different cells are sufficiently distinct from each other, *i.e.,* fingerprints serve as unique location signatures. Specifically, with the variance fingerprint definition, the subset of APs in each cell (*i.e.,* a sampled location in the radio map construction phase) that have the highest variance, across all cells with respect to their mean RSSI values, are chosen to compare with the corresponding set of APs from runtime fingerprints to find the closest matching cells.

2) *AP Visibility based:*

The visibility of APs is an important aspect for WiFi fingerprinting systems that has so far received less attention in the literature. Some proposals assume that identical set of APs are seen across the whole space of interest, whereas others implicitly suppose that the visibility of an AP is constant over time. These assumptions often do not hold in practice. To capture the impact of AP visibility on the accuracy with WiFi fingerprinting systems, we consider the following two different definitions:

Constancy: At a given location, there may be differences between different APs in terms of how often they are seen in fingerprint measurements because of weak signals, small-scale fading, beacon loss due to co-channel interference etc. The constancy definition essentially captures this aspect. Specifically, for each cell, we select those APs which appear the most number of times across multiple site survey measurements at that cell during radio map construction. The mean RSSI of this subset of APs is then compared with the runtime RSSI measurements of the same set of APs for location estimation.

Coverage: This definition captures a different spatial aspect of AP visibility. It picks, for each cell, the subset of APs that are most widely seen across all cells in the space of interest for pattern matching during location estimation.

3) Hybrid Definitions: Recall that we select a subset of APs satisfying a certain property in our alternative set of fingerprint definitions. However when using the constancy definition, we observed that often several APs are seen in a cell the same number of times. We randomly break ties with the vanilla constancy definition described above, whereas here we consider hybrid definitions that combine constancy with other similar definitions. We focus on constancy combined with either strength or stability as strength and stability show good correlation with constancy (see Table I). Based on this, we consider the following two fingerprint definitions:

Constancy+Strength: With this definition, we first rank the APs seen in a cell in the decreasing order of their constancy. Between APs with the same constancy, we prefer those with a higher strength as indicated by their mean RSSI value in the fingerprint database.

Constancy+Stability: As with the previous definition, APs seen in a cell across all measurements in the radio map construction phase are ordered based on their relative constancy so that APs with higher constancy appear earlier in the order. Then stability of the APs as defined above is used to choose among the APs with the same constancy.

TABLE I: Pearson correlation coefficient computed between constancy and strength / stability for different floors in our Forum office environment.

Floor	Constancy-Strength	Constancy-Stability
1st Floor	0.4050907	0.1883634
2nd Floor	0.6191411	0.3272367
3rd Floor	0.6430674	0.3887892
4th Floor	0.6001379	0.35482358
5th Floor	0.6507656	0.45762849

D. Location Estimation Algorithms

In our study, we consider five different location estimation algorithms. The first three belong to the deterministic techniques (*e.g.,* RADAR [1]) whereas the other two fall under the category of probabilistic techniques exemplified by Horus [2].

1) Deterministic or Nearest Neighbor (NN) Techniques: The use of nearest neighbor techniques is quite common with WiFi fingerprinting systems. Essentially, the idea is to compute the distance in signal space between pre-collected, location tagged fingerprints in a database and a runtime fingerprint to find the closest match or matches. Different NN techniques differ in the distance computation methods used. We consider three representative methods as outlined below.

Euclidean Distance: This method used in [1] and other WiFi location fingerprinting systems uses equation 1 to compute the distance between fingerprints from the database, each with an associated location and denoted by $\mathbb{S}$, with a runtime fingerprint $\mathbb{R}$. In equation 1, n is the number of APs considered in the fingerprints; in our study, this is total number of APs in the environment with the default fingerprint definition and 5 for the other definitions. And s_i is the mean RSSI value of AP i in the fingerprint from the database, whereas r_i is AP i's RSSI in the runtime fingerprint.

$$EucDist(\mathbb{S},\mathbb{R}) = \sqrt{\sum_{i=1}^{n}(s_i - r_i)^2} \qquad (1)$$

Manhattan Distance: Manhattan distance, which is also mentioned in [1], is another well-known NN method. It is defined as the sum of the absolute differences of values between fingerprint from database and runtime fingerprint as indicated by the following equation:

$$ManDist(\mathbb{S},\mathbb{R}) = \sum_{i=1}^{n}|s_i - r_i| \qquad (2)$$

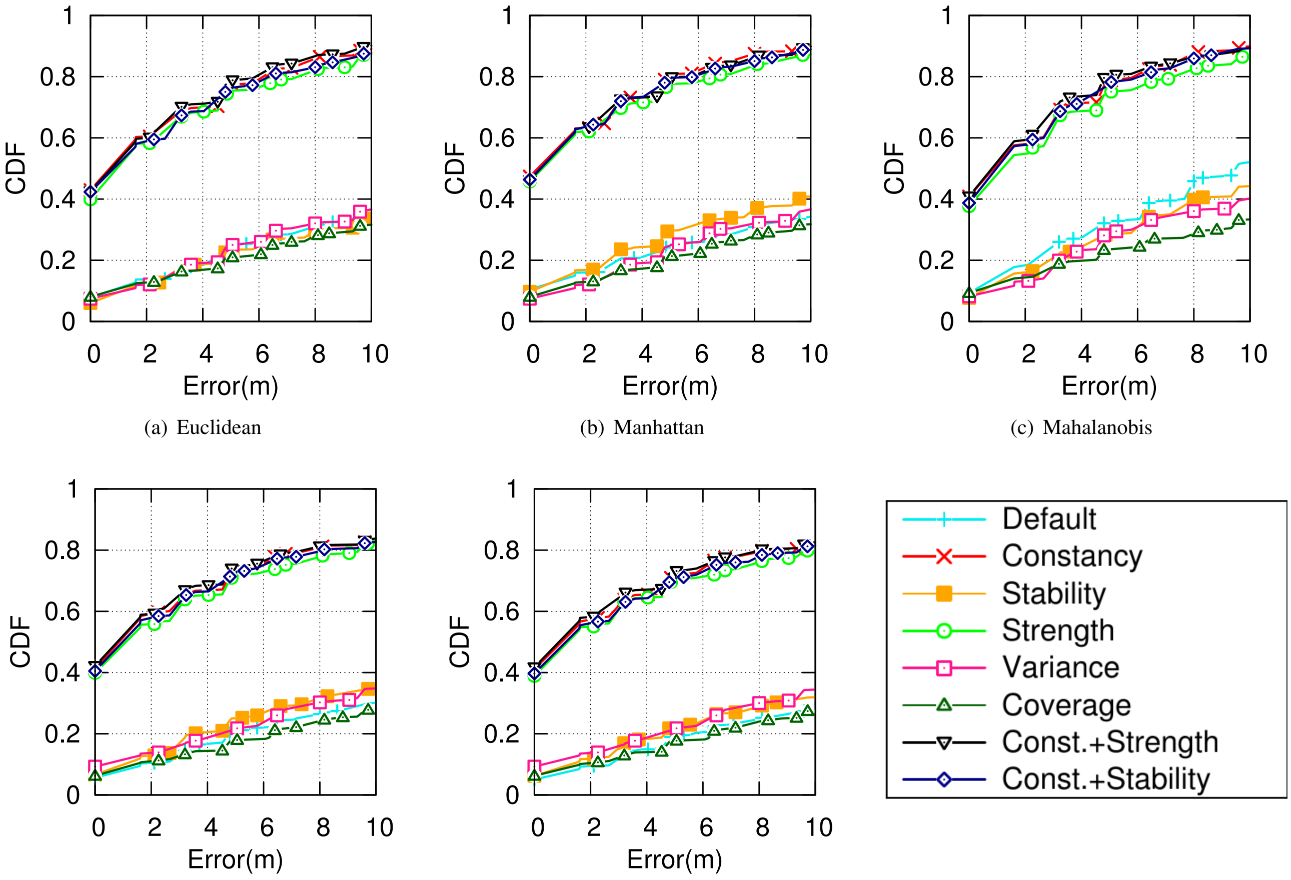
Mahalanobis Distance: Mahalanobis distance is yet another NN method considered in the WiFi fingerprinting literature (*e.g.,* see [7] and references therein). It is more sophisticated compared to the previous two methods and accounts for correlations between compared vectors. An interesting feature of Mahalanobis distance is that it is based on assumptions of stable patterns of RSSI distributions and it also takes into account variance in RSSI as done in probabilistic techniques [9], [10]. Mathematically, Mahalanobis distance computation is shown by equation 3 where S is the covariance matrix of $\mathbb{S}$ and $\mathbb{P}$ of the same distribution.

$$MahalDist(\mathbb{S},\mathbb{R}) = \sqrt{(\mathbb{S}-\mathbb{R})^T S^{-1}(\mathbb{S}-\mathbb{R})} \qquad (3)$$

2) Probabilistic Techniques: This class of techniques infer the probability that a user is at a certain location based on modeling RSSI measurements in each cell from the radio map construction phase as a probability distribution. In simple terms, they select the cell x that maximizes the conditional probability $P(x/\mathbb{R})$ given an online fingerprint $\mathbb{R}$ as the user's most likely location. Different techniques differ in the type of distribution used for RSSI modeling. We focus on two commonly considered distributions: **Gaussian** (as in [2]) and **Log-Normal**.

IV. Impact of Fingerprint Definition and Location Estimation Algorithms

In this section, we assess the relative importance of fingerprint definition in relation to location estimation algorithms for different environments. Throughout we use at least 15



(a) Euclidean

(b) Manhattan

(c) Mahalanobis

(d) Gaussian

(e) Lognormal

(f) Legend

Fig. 3: CDF of estimated location errors with different fingerprint definitions and location estimation algorithms *across all floors* in the office environment.

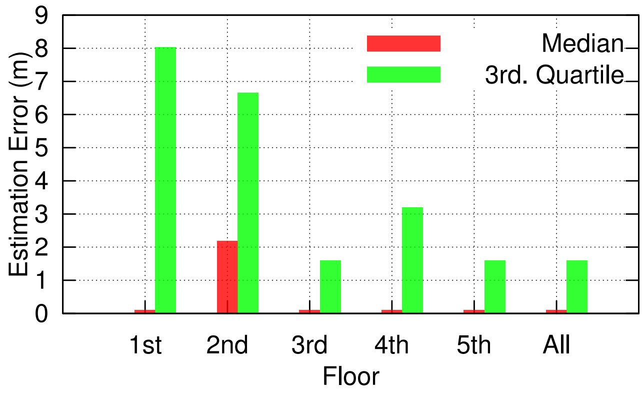


Fig. 4: Summary statistics (median and 3rd quartile) location estimation error for the best combination of fingerprint definition and location estimation algorithm for various floors separately and together in the office environment.

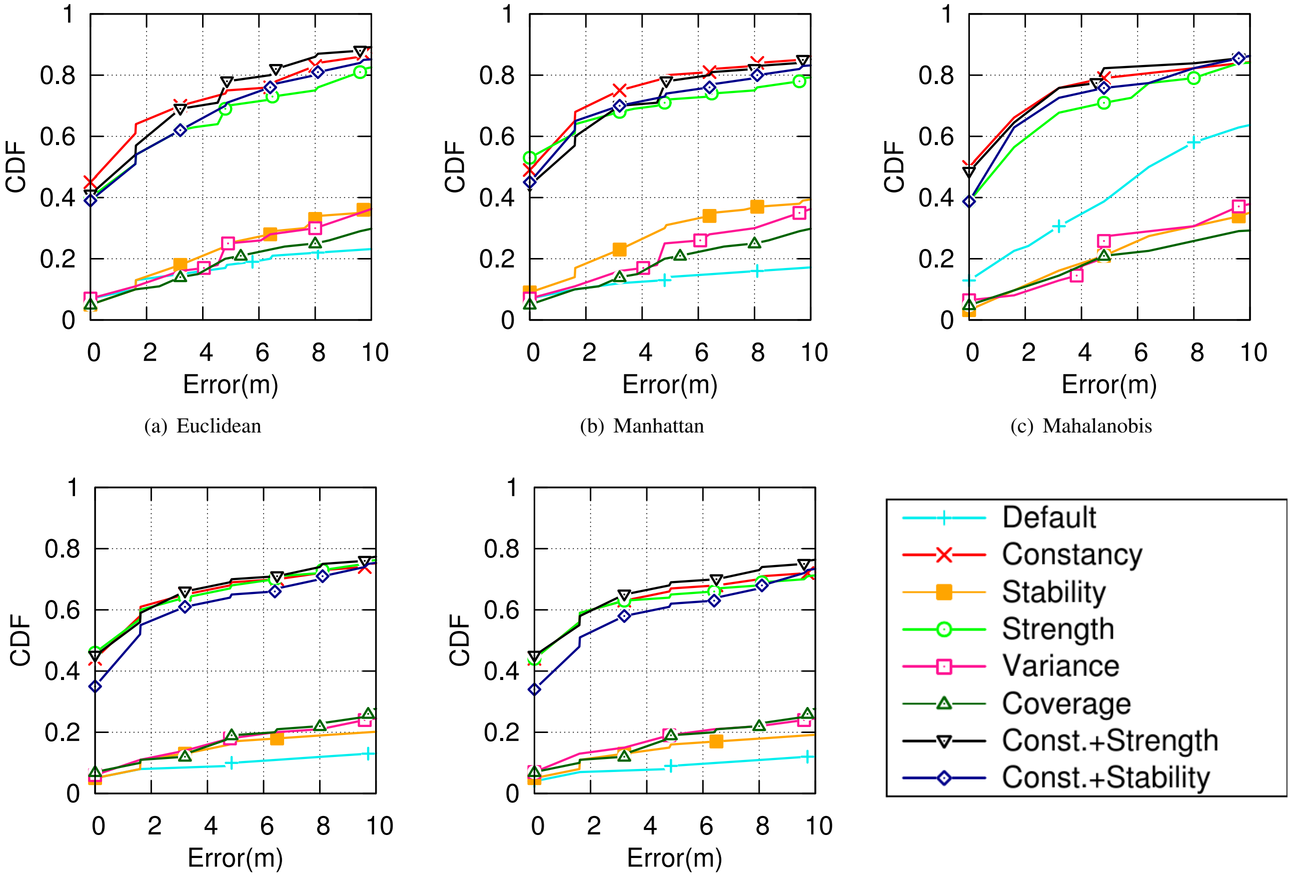
measurement samples (WiFi scans) per location for the reference fingerprint database, and 5 samples for runtime location estimation.

We look at the office environment first and then the various shopping center environments.

Office Environment. Figure 3 shows the cdf of location estimation errors with all possible combinations of fingerprint definitions and location estimation algorithms when all 5 floors in the office building are seen as one whole. We see that various fingerprint definitions appear clustered in two separate groups with significant difference in accuracy between them. Constancy, strength and the two hybrid definitions fall in the best performing group. Surprisingly, stability and variance yield poor performance for all algorithms as does coverage. As mentioned earlier in section III, default is also in the same group providing poor location accuracy.

Now turning attention to the various location estimation algorithms, we see that Manhattan distance performs slightly better among the deterministic techniques. It is noteworthy that probabilistic techniques yield poor accuracy compared to all three deterministic techniques; this is more apparent if results are compared near the right end of the plots near 10m error. We believe this is because the true RSSI distribution differs from the one chosen to model it (Gaussian or Lognormal).

Overall we can also observe that the choice of fingerprint definition has as much or more impact than the location estimation algorithm. Table II summarizes the best combina-



(a) Euclidean

(b) Manhattan

(c) Mahalanobis

(d) Gaussian

(e) Lognormal

(f) Legend

Fig. 5: CDF of estimated location errors with different fingerprint definitions and location estimation algorithms for *first floor* in the office environment.

tion of fingerprint definition and location estimation algorithm which turns out to be Strength with Manhattan distance for the whole building case. The best combination is obtained by first identifying the combination providing least median estimation error; in case there are several such combinations then their performance is compared in terms of 3rd quartile estimation errors; if there are still multiple candidates then the one providing the smallest maximum error is chosen as the best combination.

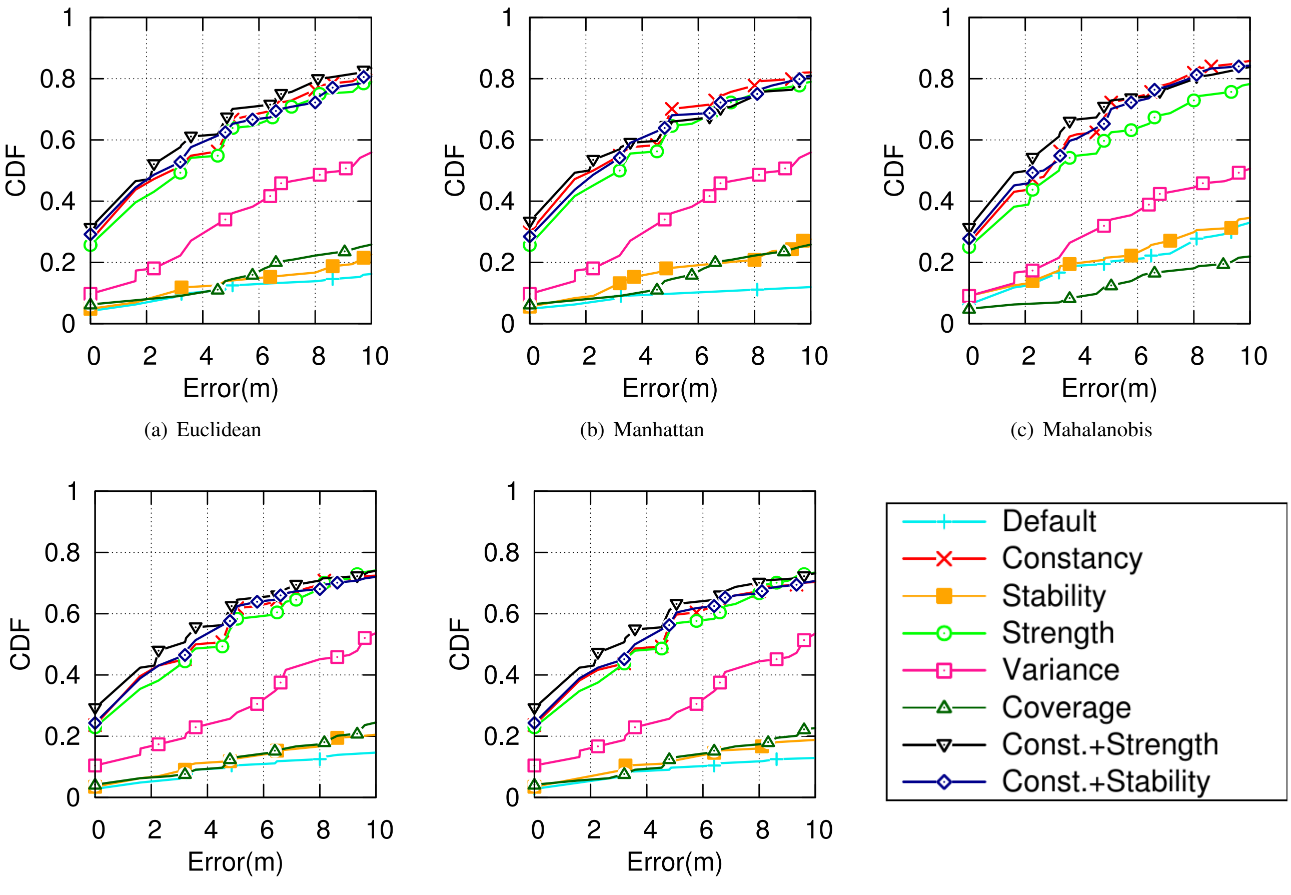
When each floor is seen in isolation, Table II also shows that the best combination is different between floors. This is also evident when we look at the median and 3rd quartile estimation errors in Figure 4. We see that the second floor has higher errors. This is because of the open area on that floor where all combinations have difficulty telling apart different cells within that open area. CDFs of location estimation errors for the first and second floors shown in Figure 5 and Figure 6, respectively, further illustrate this point. We also notice that differences between different fingerprint definitions and location estimation algorithms become more apparent at the individual floor level.

Shopping Centres. Different shopping centers are quite different in terms of their location estimation error statistics as shown in Figure 7. We can see that shopping center 3 is the easier of the three to localize as it is more compact and rich in multipath.

TABLE II: Office Environment: best combination of fingerprint definition and location estimation algorithm

Floor	Loc. Est. Algo	Fingerprint Defn.
1	Manhattan	Strength
2	Mahalanobis	Constancy+Strength
3	Manhattan	Constancy
4	Manhattan	Constancy+Stability
5	Manhattan	Strength
All	Manhattan	Strength

Notice also that errors in Figure 7 are also higher compared to Figure 4, partly because of the sparser location sampling in the former as mentioned in section III. As with the office environment, we see from Table III that best combination changes from one environment to the other. This is true even between floors within shopping center 3, the only one spanning 2 floors in our study. But interestingly, Mahalanobis distance always emerges as the location estimation algorithm in all best combinations cases.



(a) Euclidean

(b) Manhattan

(c) Mahalanobis

(d) Gaussian

(e) Lognormal

(f) Legend

Fig. 6: CDF of estimated location errors with different fingerprint definitions and location estimation algorithms for *second floor* in the office environment.

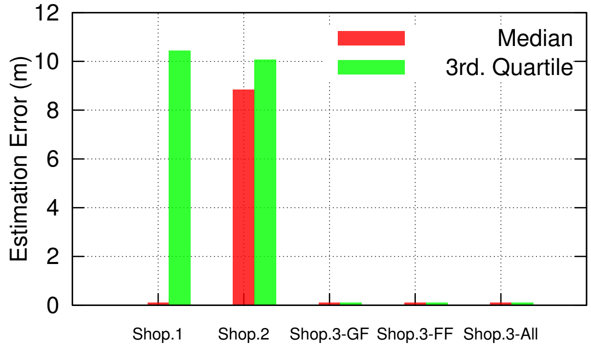


Fig. 7: Summary statistics (median and 3rd quartile) location estimation error for the best combination of fingerprint definition and location estimation algorithm for different shopping center environments.

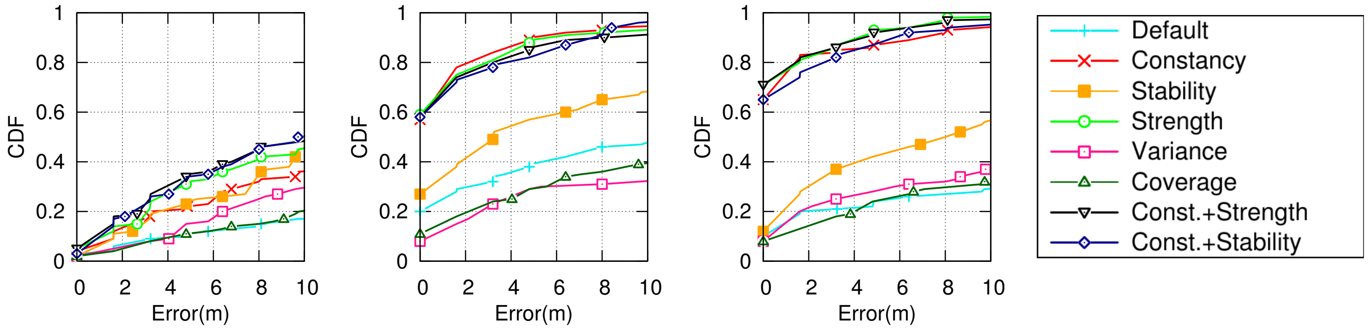
V. The Impact of Frequency Band

In this section, we explore the impact of frequency band (2.4GHz vs. 5GHz) on WiFi fingerprinting accuracy. While 2.4GHz was the only band originally used for WiFi, increasingly 5GHz is also being used despite its relatively poorer propagation characteristics resulting from higher frequency operation. This is because 5GHz band is less crowded and also there is far more spectrum available in 5GHz band. From a WiFi fingerprinting system perspective, in a typical environment today with APs using both 2.4GHz and 5GHz bands, a measurement sample (WiFi scan) obtained either during the radio map construction phase or subsequent runtime phase will likely include a mix of 2.4GHz and 5GHz APs. This in turn could impact the accuracy of the WiFi fingerprinting system as signals from these two bands behave differently.

TABLE III: Shopping centers: best combination of fingerprint definition and location estimation algorithm

Environment	**Loc. Est. Algo**	**Fingerprint Defn.**
Shop. Ctr. 1	Mahalanobis	Stability
Shop. Ctr. 2	Mahalanobis	Constancy+Stability
Shop. Ctr. 3-GF	Mahalanobis	Constancy
Shop. Ctr. 3-FF	Mahalanobis	Constancy+Stability
Shop. Ctr. All	Mahalanobis	Constancy

To study the impact of frequency band on WiFi fingerprinting, we used a smart phone that supports both 2.4GHz and 5GHz bands (Samsung Galaxy S3) to collect multiple samples for each measurement location shown in Figure 1(a) for the first floor of the Forum office environment.



(a) 2.4GHz

(b) 5GHz

(c) Both 2.4GHz and 5GHz

(d) Legend

Fig. 8: CDF of estimated location errors across 2.4GHz and 5GHz bands and together for different fingerprint definitions and Euclidean distance method for the first floor in the office environment.

Figure 8 shows the CDF of location estimation errors for the cases where only APs from one band are considered as well as the case considering APs from both bands. We show results for only one location estimation algorithm (Euclidean distance) for brevity as the results are qualitatively similar for other algorithms. Results in Figure 8 show that the cases including APs from 5GHz band show a clear and significant benefit compared to using only the 2.4GHz band even though the number of APs in the environment are evenly distributed across the two bands.

To better understand the reasons behind the improvement in WiFi fingerprinting accuracy obtained using 5GHz band, we setup an AP with a multiband WiFi card and had a client in the form of laptop with AirPcap USB dongle[2] listening to beacons sent from the AP on channels from both bands. Figure 9 shows the mean and standard deviation of RSSI of AP beacons, separately for each band. While the lower mean RSSI in the 5GHz is expected, the relatively higher standard deviation in RSSI in 2.4GHz is interesting and we believe is also the key reason why using APs for 2.4GHz band alone results in poor location accuracy. We also conducted a similar experiment in two shopping centers using a AirPcap equipped laptop listening to beacons from already existing multiband APs for 1.5 hours and find that beacons received on 2.4GHz consistently show greater variation in RSSI.

From inspecting the packet logs in the above experiments, we find that beacons in 2.4GHz are transmitted at 1Mbps 802.11b DSSS bit-rate, whereas 5GHz beacons are sent at OFDM based 6Mbps bit-rate. This difference may explain the high variation in RSSI seen for beacons on 2.4GHz. Note that RSSI is measured only for the PLCP header of received frames. The 48 bits long PLCP header for DSSS 1Mbps BPSK modulation takes 48us to transmit whereas the same length PLCP header takes only 4us at OFDM 6Mbps rate. The shorter duration for RSSI sampling in 5GHz makes it relatively less affected by temporal signal variations due to people movement etc., thereby resulting in a more stable RSSI.

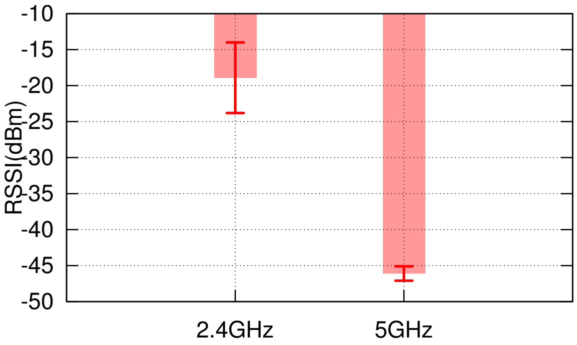


Fig. 9: Mean and standard deviation RSSI of beacons received on 2.4GHz and 5GHz bands from the same AP.

TABLE IV: Best combination of location estimation algorithm and fingerprint definition including and excluding VAPs.

Case	Loc. Est. Algo	Fingerprint Defn.
Including VAPs	Manhattan	Strength
Excluding VAPs	Mahalanobis	Constancy+Strength

We also carefully examined whether low RSSI variation in 5GHz is due to low co-channel interference. Towards this end, we setup an AP transmitting beacons in a channel of 5GHz band and an interfering node (on the same channel) with a modified device driver with CCA (Clear Channel Assessment) disabled so that it can continuously transmit without regard to whether channel is idle or busy. By measuring loss and signal strength of beacons at a client station associated with the AP, we find that increase in traffic intensity from the interfering node only increases the beacon loss but does not affect RSSI.

We have also obtained similar qualitative results comparing different bands for shopping centers but we do not include them due to space limitations.

VI. The Effect of Virtual Access Points

In this section, we study, again for the first time in the literature, the effect of virtual access points (VAPs) on WiFi fingerprinting accuracy. VAP is a way to realize multiple APs, each potentially using a different security mechanism and targeting a different set of users, with a single physical AP via time sharing. It is the wireless counterpart of VLANs. The

[2] http://www.metageek.net/products/airpcap/

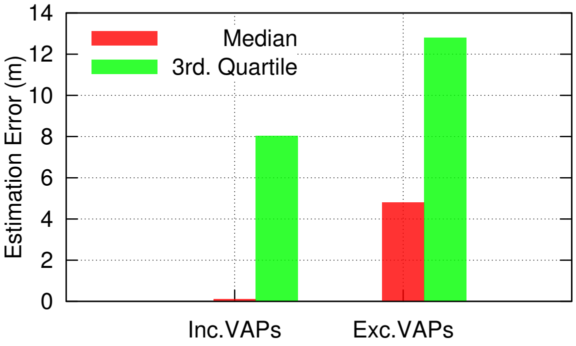


Fig. 10: Summary statistics (median and 3rd quartile) location estimation error for the best combination of fingerprint definition and location estimation algorithm with VAPs included and excluded for the first floor of the office environment.

BSSIDs of VAPs corresponding to a physical AP are typically derived from the BSSID (MAC address) of the physical AP. From our study of WLAN deployments in offices and public spaces, we observe that VAPs are common today.

Our interest here is to study the impact of the presence/absence of VAPs on WiFi fingerprinting. Towards this end, we studied the effects of VAP presence of both office and shopping center environments. However for the sake of brevity, we focus on the results for the first floor of the Forum office environment. As noted earlier in section III.B, each of the physical APs in the university WLAN network advertise two VAPs. On the first floor there are 33 university run APs resulting in 66 VAPs, plus 10 other non-VAP APs. Thus in total there are 76 APs in total when VAPs are counted, and 43 otherwise. In this environment we find that BSSIDs of VAPs share the first ten digits with the BSSID of their corresponding physical AP. It is relevant for WiFi fingerprinting to understand how the beacons of VAPs are transmitted. By capturing all beacons in the air with a laptop running Kismet application, we find that beacons for each of the VAPs corresponding to a physical APs are sent within a short period of 100ms, the default beacon transmission interval. This suggests all VAPs can be usually detected via passive scanning as the time spent on a channel before hopping to another channel is 100ms by default.

To study the effect of VAPs, we consider two cases, one with VAPs included and the other in which VAPs are excluded. The case with VAPs included simply treats each VAP as a separate physical AP; this is what we did so far in this paper. In contrast, only one VAP per physical AP is retained in the latter case. Figure 10 differentiates between these two cases in terms of their median and 3rd quartile errors considering the best combination of fingerprint definition and location estimation algorithm for each case (see Table IV). Clearly, including VAPs significantly reduces location estimation error, especially in terms of median. Figure 13 demonstrates the benefit from considering VAPs in more detail.

We attribute the gain seen from including VAPs to two reasons. Firstly, including VAPs increases the AP density which tends to have a positive correlation with higher location accuracy for WiFi fingerprinting systems. For the results shown here, the case with including VAPs has 76 APs in total whereas excluding VAPs brings that down to 43, both for the same area. Secondly, even though we may expect VAPs corresponding to a physical AP to have identical signal characteristics, this is not always the case as beacons from different VAPs are separated in time each capturing a slightly different time-varying environment context as demonstrated by Figure 11 and Figure 12.

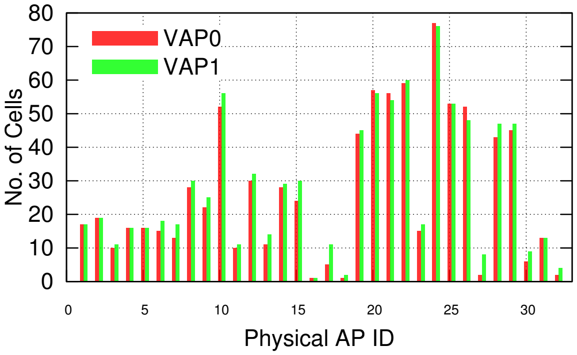


Fig. 11: Relative differences in signal coverage between each pair of VAPs corresponding to a physical AP in terms of cells where they are seen.

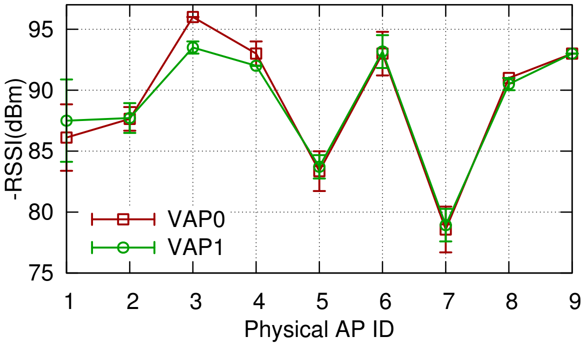
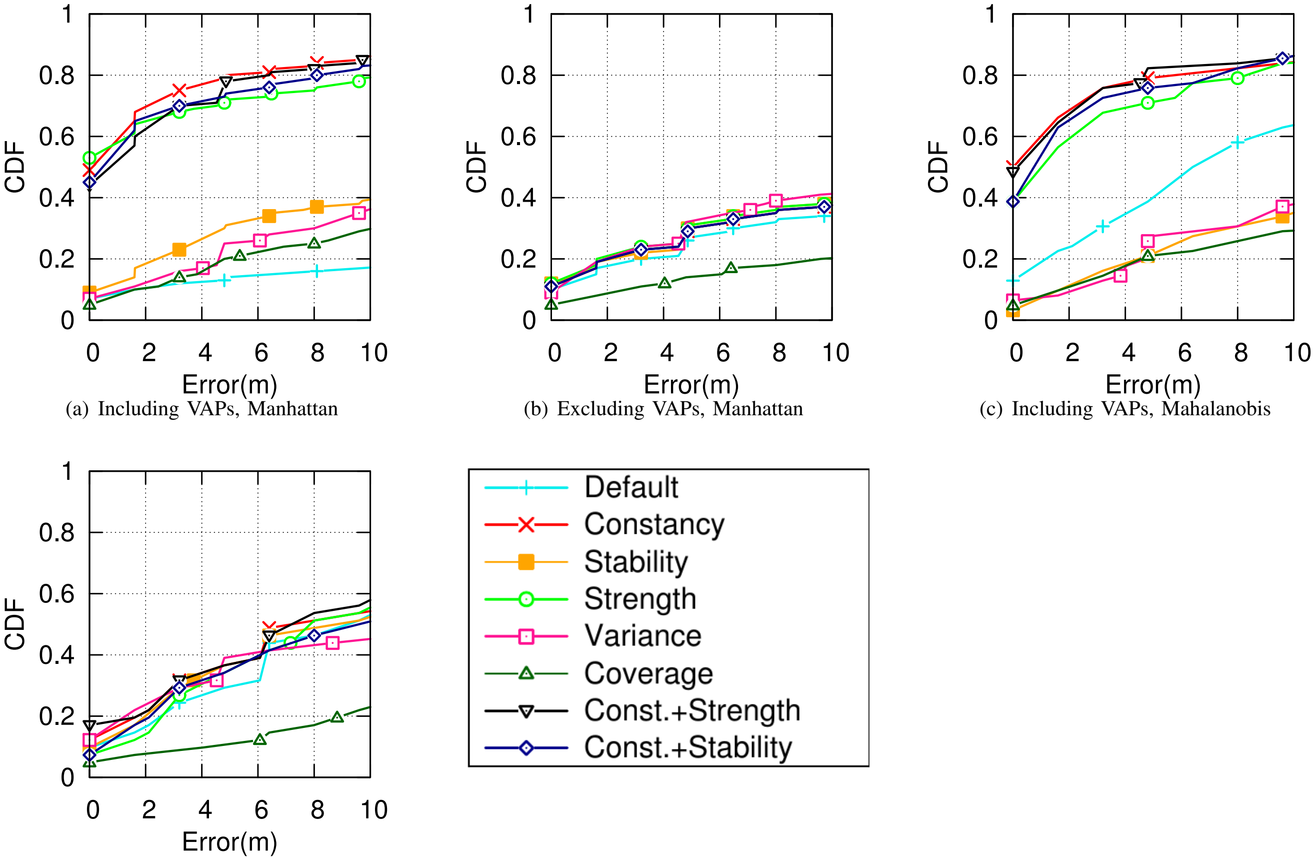


Fig. 12: Differences in mean and standard deviation of RSSI of each pair of VAPs as seen from a cell that shows maximum improvement in location accuracy from including VAPs.

VII. Conclusions

We have examined the impact of fingerprint definitions along with location estimation algorithms on WiFi fingerprinting location accuracy across diverse environments. We find that the combination of fingerprint definition and location estimation algorithm that yields best location accuracy is highly dependent on the environment and even specific floor within a given environment. We also find that the choice of frequency band (2.4GHz vs. 5GHz) and inclusion of VAPs has a significant impact on the location accuracy of WiFi fingerprinting systems; we analyze the potential reasons to explain these findings.



(a) Including VAPs, Manhattan

(b) Excluding VAPs, Manhattan

(c) Including VAPs, Mahalanobis

(d) Excluding VAPs, Mahalanobis

(e) Legend

Fig. 13: CDF of estimated location errors including and excluding VAPs for different fingerprint definitions and Manhattan/Mahalanobis distance methods for first floor in the office environment.

References

[1] P. Bahl and V. Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System. In *Proc. IEEE INFOCOM*, 2000.

[2] M. Youssef and A. Agrawala. The Horus Location Determination System. *Wireless Networks*, 14(3), 2008.

[3] P. Bolliger. Redpin – Adaptive, Zero-Configuration Indoor Localization through User Collaboration. In *Proc. ACM MobiCom MELT Workshop*, 2008.

[4] J. Park et al. Growing an Organic Indoor Location System. In *Proc. MobiSys*, 2010.

[5] A. Rai et al. Zee: Zero-Effort Crowdsourcing for Indoor Localization. In *Proc. ACM MobiCom*, 2012.

[6] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche. A Comparative Survey of WLAN Location Fingerprinting Methods. In *Proc. 6th Workshop on Positioning, Navigation and Communication (WPNC'09)*, 2009.

[7] K. Kaemarungsia and P. Krishnamurthy. Analysis of WLANs Received Signal Strength Indication for Indoor Location Fingerprinting. *Pervasive and Mobile Computing*, 8(2), 2012.

[8] G. Lui, T. Gallagher, B. Li, A. G. Dempster, and C. Rizos. Differences in RSSI Readings Made by Different Wi-Fi Chipsets: A Limitation of WLAN Localization. In *Proc. International Conference on Localization and GNSS (ICL-GNSS)*, 2011.

[9] P. Mahalanobis. On the Generalised Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 1936.

[10] H. Shin and H. Cha. Wi-Fi Fingerprint-Based Topological Map Building for Indoor User Tracking. In *Proc. 16th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 2010.