WiFi Fingerprint Localization in Open Space

Banghui Lu†∗, Jianwei Niu∗, Junghyun Jun†, Long Cheng†∗, Yu Gu†∗†
∗State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China
†Singapore University of Technology and Design, Singapore

Abstract—In recent years, WiFi fingerprint-based localization has received much attention due to its deployment practicability. In this work, we investigate the impact of different indoor environmental factors on the performance of WiFi-fingerprinting. We find that, WiFi-fingerprinting highly depends on the deployed indoor environment. In an open space, it is quite challenging to find spatially varying but temporally stable signatures for adjacent reference locations. To address this issue, we propose a heatmap-based WiFi fingerprinting method (called HMF) by utilizing a period of history location estimations as an additional input to improve WiFi fingerprint localization in open space environment. Our experimental results show, HMF can improve existing WiFi fingerprinting schemes like Radar and Horus by 28% and 80% in moderately open space like wide corridor.

I. INTRODUCTION

With the increased deployments of ubiquitous 802.11 WiFi access points (APs), WiFi fingerprint-based indoor localization is considered as the most promising localization method due to its deployment practicability.

A WiFi-fingerprinting system works in two phases: an offline training phase and an online localization phase. The offline phase collects radio signal strength (RSS) measurements of WiFi APs (known as fingerprint map) from different pre-known locations and stores them to a database as the training set. In the online phase, users infer their current location based on the observed RSS measurements, through finding the closest match in the database.

Several existing research demonstrated that WiFi-fingerprint based localization can achieve average localization accuracy of less than 10m. For example, Horus [1], statistically models WiFi signal strength observations to estimate the user location, demonstrated that it can achieve sub-meter accuracy. EZ [2] models the WiFi signal based on well known wireless communication model and demonstrates that they can maintain localization error within 2-7m. Radar [3] demonstrated that their system can achieve average localization accuracy below 3m. ZEE [4] and EZ [2] show that the accuracy of Horus was 3m and 4m under their experimental testbeds, respectively, rather than the sub-meter accuracy achieved in the original Horus [1]. These results imply that the performance of WiFi-fingerprinting depends on the indoor environment in which it is deployed.

In this paper, we first present empirical evaluation of WiFi-fingerprint based localization in four different indoor environments: large open space, small open space, wide corridor, narrow corridor. Our empirical experiment results reveal two weakness in the design of WiFi-fingerprint based localizations: (1) Its performs is limited in a large open space environment. (2) Its accuracy decrease quickly overtime unless the fingerprint map is frequently updated with newly collected WiFi radio signals at every locations.

In summary, our contributions are followings:

• We evaluate the WiFi fingerprinting localization accuracy at different environments and show how accuracy changes overtime without updating the fingerprint radio map.
• To the best of our knowledge, this is first work analyzing the performance of WiFi fingerprint radio map based on a new metric γ which measures the spaciousness of the indoor environment in which WiFi fingerprinting is deployed.
• We propose a heatmap-based WiFi fingerprinting (HMF) by utilizing a period of history location estimations as an additional input. Experimental results demonstrate that our method can improve existing Radar and Horus by an average of 28% and 80% in moderately open space like wide corridor.

II. EMPIRICAL EVALUATION

In this section, we empirically evaluate the impact of different indoor environments and fingerprint variation overtime on the localization accuracy.

A. Impact of Indoor Environment

Fig. 2. WiFi fingerprinting accuracy over time

In order to generalize the impact of indoor environment, we define spatial similarity γ, which is the average euclidean distance between every pair of reference points in the WiFi-fingerprint radio map, if a pair of reference points has a line-of-sight and their distance is less than maximum transmission range of AP. The metric γ attempts to measure the spaciousness of an area relative to WiFi-fingerprint localization methods. For example, the value of γ would be almost zero for an office building with many small rooms if reference points are located in rooms. In contrast to office building, the γ would be reasonably big for an open area like a large concert hall with too many reference points have line-of-sight.
We have conducted WiFi-fingerprinting experiments at four distinctive areas: office building (narrow corridor), lecture theaters (small open space), wide corridor, complete open space. On the average, each location is covered by 12 access points. At each location, we collect 1000 fingerprints continuously. The test data are collected by different phones within several days after training data was collected.

Figure 1 shows the scenarios of office building, lecture theaters, wide corridor, and complete open space. Office building has many small rooms with long and narrow corridors. Lecture theaters is an area where class rooms of size greater than 20m² are clustered together. Wide corridor is a 107 × 7 meters long and wide area. The most of reference points in this area are deployed on straight wide corridor (average width is 5m) with some rooms on each side of the corridor. Complete open space is spacious and complete open which area is greater than 200m². The reference points in complete open space are uniformly deployed in these test beds.

For office building testbed, no pair of reference points has line-of-sight view from each other and therefore its \( \gamma = 0 \). The \( \gamma \) for other three testbeds are shown in the Table I.

<table>
<thead>
<tr>
<th>Area Type</th>
<th>Office building</th>
<th>Lecture theaters</th>
<th>Wide corridor</th>
<th>Complete open space</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0.0</td>
<td>6.8</td>
<td>7.3</td>
<td>9.7</td>
</tr>
</tbody>
</table>

**TABLE I**

This table shows \( \gamma \)'s for different areas

Figure 2 shows WiFi fingerprint accuracy at four indoor areas with different \( \gamma \) levels. The office building with \( \gamma = 0 \) has the lowest localization error of 18%. Whereas the highest localization error of 73% is the complete open space in which every pair of reference points is in line-of-sight. Also the results in Figure 2 indicates that there exists a proportional relationship between WiFi fingerprinting localization error probability and the \( \gamma \). Therefore, one can use this result to predict whether WiFi fingerprinting would be a sufficient localization solution for their indoor environment. This result in Figure 2 also demonstrates that WiFi fingerprinting would be an ideal solution for logical localization at a room level but not sufficient for physical localization requires the accuracy of less than 10m in an complete open space where \( \gamma \) is high.

**B. Impact of Fingerprint Variation Overtime**

The indoor environment changes dynamically overtime. Due to this dynamic changes, reliability of fingerprint map expected to deteriorate overtime. In order to understand how the accuracy of WiFi fingerprinting changes overtime, we have collected the 1000 fingerprints everyday at every reference points over a period of one month.

Figure 3 represents the changes in the probability of localizing users with accuracy below 2m or below 9m with respect to time (in the scales of days). Figure 3 shows that the probability of maintaining the localization accuracy below 2m exponentially decreases while the probability of maintaining the localization accuracy below 9m remains reasonably stable overtime. This indicates the reliability of WiFi-fingerprint based localization would be low for applications which require high accuracy (< 2m) in the complete open space.

**III. HEATMAP-BASED FINGERPRINTING**

We have empirically demonstrated in Section II that the expected performance of WiFi fingerprinting strongly depends on the spatial similarity and temporal variation between any nearby reference points. This suggests that additional input is necessary in order to guarantee high accuracy of WiFi fingerprinting method for the indoor environment with high \( \gamma \). Here, we propose heatmap-based fingerprinting, a quick and effective solution to improve RSS distribution based WiFi fingerprinting method by utilizing history location estimations.

**A. Motivation of Heatmap Design**

RSS probability distribution based WiFi fingerprint localization has been performed in some previous works [1], [8]. Given the observed RSS vector \( \vec{S} = (s_1, ..., s_k) \), while \( s_k \) is the measured RSS of number \( k \) AP. The predicted location \( l^* \) was computed in Equation 1.

\[
l^* = \arg \max_l P(l|\vec{S}) = \arg \max_l \frac{P(\vec{S}|l) \ast P(l)}{P(\vec{S})} \tag{1}
\]

Both of these two works have two assumptions. Firstly, they assumed that all the potential locations are equally likely and considered the term \( P(\vec{S}) \) as a constant which is factored out from Equation 2. Secondly, they consider \( P(\vec{S}) \) remains stable for long duration of time which is also a constant. So actually the predicted location \( l^* \) was computed by Equation 2.

\[
l^* = \arg \max_l P(l|\vec{S}) = \arg \max_l \prod_{i=1}^{k} P(s_i|l) \tag{2}
\]

\( P(s_i|l) \) is estimated from the histograms which contains the RSS distributions.
B. Heatmap-based WiFi Fingerprinting

The uniform distribution of user’s location over the set of possible locations is a common assumption of WLAN localization systems [1]. However, if we know the distribution of user’s location, that is, user has higher probability at a certain location or area, it can help to increase the localization accuracy. Many previous works, such as Horus [1], assume that \( P(l) \), the probability of user’s current location, is uniformly distributed over all reference points. But, in fact due to movement restriction \( P(l) \) is not uniformly distributed and changes over time. So, instead of set \( P[l] \) and \( P[S] \) to a constant, we exploit the distribution of user’s location \( P(l) \) by heatmap, which is calculated by utilizing a period of history location estimations. In section III-C, we show how the heatmap was built. Let \( L \) be the list of all reference locations and \( l_i \) be \( i_{th} \) reference location in \( L \). The heatmap of time \( t \) can be described by \( H_t = \{ P[l_1], P[l_2], P[l_3], ..., P[l_n], P[l_n] \} \). \( H_t \) represents the distribution of user’s potential location at time \( t \). For our HMF, the estimated location \( l^* \) was calculated by Equation 3.

\[
l^* = \arg\max \frac{P[S[l]] \ast P[l]}{\sum_{i=1}^{n} P[S[l_i] \ast P[l_i]} \quad (3)
\]

We introduce the bounding \( P(l) \) to further improve the performance of WiFi-fingerprint based localization.

C. Fingerprint Heatmap Building

The instantaneous location estimation of fingerprint based localization is not reliable. However, since the basic WiFi fingerprint localization has a reasonable accuracy, based on the observation of a certain period, we can roughly know which area or location has higher probability to be the ground truth location. Here, we use the connectivity between reference points and statistics from past estimated locations to build the heatmap. For example shown in Figure 4, two reference points A and B are connected since these two locations are physically connected and there is no other reference points between them. Reference point A and C is not directly connected since we must come close to reference point B before we can reach C. The probability of each reference point is assigned by adding a set of consecutive WiFi fingerprint estimation. For example, heatmap shows reference point A and B have probability of 0.8 and 0.2, respectively and then user start to move from A to B. If live WiFi fingerprint estimates \( P(S|B) = 0.4 \) and \( P(S|D) = 0.6 \), we update \( P(B) \) by \( P(S|B) + 0.6 \) and normalize all the probabilities in heatmap, but ignore \( P(S|D) \) since D is not included in heatmap. The heatmap indicates that the current potential locations of users are A and B, but without heat map, D will also be a potential location with high probability. The detail algorithm is shown in heatmap building algorithm 1.

Algorithm 1: Heatmap building

**Input:** fingerprint map \( F \), reference location set \( L \), live WiFi fingerprint \( S_i \) at time \( t \), heatmap sliding window \( \omega \) and \( \delta \) for heatmap size reduce.

**Output:** \( H_t \) which represent the heatmap of time \( t \).

\[
\forall l \in L, \ P[l_i] = 1/\text{size}(L);
\]

while \( t_{now} < t_{end} \) do

\[
given \ S_{t_{now}} \text{, compute } V_{t_{now}} = \{ P[S_{t_{now}} | l_i] \mid l_i \in \mathbb{L} \};
\]

for \( l_i \in \mathbb{L} \) do

\[
P[l_i] = \frac{1}{\omega} \sum_{t_{now} - \omega}^{t_{now}} P[S_i | l_i], \ \forall l \in \mathbb{L};
\]

\[
H_{t_{now}}[l] = P[l_i];
\]

for \( l_i \in \mathbb{L} \) do

\[
\text{if } H_{t_{now}} < \delta \text{ then } H_{t_{now}} = 0;
\]

normalize \( H_t \);

save \( V_{t_{now}} \) to Queue \( Q \);

for \( V_t \in Q \) do

\[
\text{if } t < t_{now} - \omega \text{ then } \text{remove } V_t \text{ from } Q;
\]

return \( H_t \);

IV. HMF Evaluation

In this section, we present the experimental result of HMF localization and compared it with Horus and Radar which are two well accepted WiFi fingerprint solutions in practice.

A. Experiment Setup

1) Testbed: We evaluate our HMF localization scheme under two different testbeds: wide corridor and complete open area. The floor layout of these two testbeds are shown in Figure 1. The dimension of the wide corridor testbed is \( 107m \times 7m \) and there are totally 51 APs were detected in this test bed. Another one is the open space testbed and it’s size is \( 36m \times 24m \). In this testbed, there are totally 26 APs. We have a total of 89 reference points in wide corridor testbed. The average distance between these reference points are 2.5m. On the average, each location is covered by 12 access points. For test bed 2, there are 77 reference locations and each location can detect about 15 APs.

2) Offline Data Collection: We use 10 Sony Acro S smart phones to build the Fingerprint training set and perform the realtime experiment. WiFi training data that was used to build WiFi fingerprint map was collected at different time scales of days, weeks, and months, respectively. At each time scale, we collect WiFi training data of 1000 fingerprints continuously at each reference point.
3) Realtime Experiment: In order to evaluate the expected performance of HMF in practice, we walk through all the tags one by one. At each reference location, we rest for several seconds and then move ahead.

B. Experimental Results

Figure 5 depicts the CDF of localization error observed by Radar and Horus at the wide corridor testbed. Our result shows Radar performs better than Horus over all localization error range. In the same figure, we show that utilizing heatmap improves average performance of normal Radar by 28% and 5% with probability of 0.8. Utilizing heatmap on Horus improves average performance of Horus by 80% and 11% with probability of 0.8.

Figure 6 depicts the distribution of localization error observed by Radar and Horus in the complete open space testbed. Our result shows Radar performs better than Horus over all localization error range. The CDF looks more like a step function. This is due to a fact that reference points are deployed uniformly in the complete open space testbed and the distance between every pair of any adjacent point is 3m. In the same figure, we show that utilizing heatmap improves average performance of normal Radar by 10% and 41% with probability of 0.8. Utilizing heatmap on Horus improves average performance of Horus by 33% and 41% with probability of 0.8. Our experiment result in Figure 7 show that resting time of 4s is enough to localize the user with below 2m accuracy.

Notice that the improvement is less significant in this testbed compared to wide corridor. This is due to heatmap being able to utilize some unique reference points which has very low spatial similarity with its nearby reference points in the wide corridor whereas there is almost no unique reference points in complete open space testbed.

Figure 8 shows the relation between training database size and localization accuracy. When training size of Horus increases from 300 WiFi fingerprint scanning to 1000 the average localization error of Horus improves by 9%. In contrast to regular Horus, heatmap-based Horus do not show clear improvement when training size increase. This means employing heatmap can reduce the dependency of WiFi fingerprinting to its training set size.

V. CONCLUSION

In this work, we investigated the impacts of spatial and temporal similarity by conducting extensive experiments under 4 different testbeds in which their spatial similarities are different. Our initial result shows that the spatial similarity is high in case of large open space. Consequently, WiFi fingerprint performance reduces in those open environments. We also show RSS variations over time can reduce the reliability of WiFi fingerprinting radio map to 40% only after one day if location accuracy of less than 2m is required. These two results demonstrate that WiFi fingerprinting alone is not sufficient and additional information must be presented as an additional input to WiFi fingerprinting to achieve higher localization accuracy in the open space.

Based on the empirical results, we proposed heatmap-based WiFi fingerprinting, fast and effective method improving WiFi fingerprinting by utilizing the information from the history location estimations. Our performance evaluation indicates that heatmap-based WiFi fingerprinting is effective and improves the localization accuracy by 40% for all four different testbeds compared to regular Horus and Radar. It also maintains its accuracy for longer duration of time.

ACKNOWLEDGEMENTS

This work was supported by the Research Fund of the State Key Laboratory of Software Development Environment under Grant No. BUAA SKLSDE-2012ZX-17, the National Natural Science Foundation of China under Grant No. 61170296 and 61190125, the Program for New Century Excellent Talents in University under Grant No. NECT-09-0028, CPSF 2013MS30511, Singapore-MIT IDI IDD61000102a, IDG31100106a, and NRF2012EWT-EIRP002-045.

REFERENCES