2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress

A Graph-based Method for Indoor Subarea Localization with Zero-configuration

Yuanyi Chen[1*], Minyi Guo[1†], Jiaxing Shen[2], and Jiannong Cao[2]
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University
Email: *cyyxz@sjtu.edu.cn, †guo-my@cs.sjtu.edu.cn
[2]Department of Computing, The Hong Kong Polytechnic University, Email: {csjshen, csjcao}@comp.polyu.edu.hk

***Abstract*—Indoor subarea localization remains an open problem due to existing studies face two main bottlenecks, one is fingerprint-based methods require time-consuming site survey and another is triangulation-based methods is lack of scalability in large-scale environment. In this paper, we aim to present a graph-based method for indoor subarea localization with zero-configuration, which can be directly employed without offline manually constructing fingerprint map or pre-installing additional infrastructure. To accomplish this, we first utilize two unexploited characteristics of WiFi radio signal strength to generate logical floor graph, and then formulate the problem of constructing fingerprint map in terms of a graph isomorphism problem between logical floor graph and physical floor graph. Then, a Bayesian-based approach is utilized to estimate the unknown subarea in online localization. The proposed method has been implemented in a real-world shopping mall and extensive experimental results show that our method can achieve competitive performance comparing with existing methods.**

I. Introduction

Recent years have witnessed an increasing attempt on indoor subarea localization in view of its importance to many location-based services, such as indoor advertising [1], patient activity monitoring [2] and indoor check-in services [3]. Since traditional GPS positioning technique is infeasible in indoor environment and the positioning accuracy of cellular-based method is not enough, localization methods based on radio signal strength (RSS) has attracted enormous research from both academia and industry. Existing RSS-based localization methods either require time-consuming site survey or huge costs for deploying additional hardware. Therefore, indoor subarea localization remains an unsolved problem according to the report from Microsoft indoor localization competition [4]. In general, existing RSS-based localization methods can be divided into two categories: infrastructure-based methods and infrastructure-free methods.

Infrastructure-based methods require pre-installed hardware for localization, such as UWB [5], ZigBee [6] or wearable sensor [7], [8], which make this kind of system unsalable to large-scale environment. To address this drawback, many infrastructure-free localization systems [9]–[11] without requiring additional hardware have been proposed. One of the most promising methods is using WiFi RSS, which is mainly attributed to the widespread deployment of WLAN infrastructure.

Previous localization methods using WiFi RSS include geometric-based scheme and fingerprint-based scheme. Geometric-based scheme utilizes geometry relation between the unknown location and more than two reference locations for localization, such as TOA [12], TDOA [13] and AOA [14]. Geometric-based scheme requires prior knowledge of WiFi access point (AP) and radio signal propagation model in indoor environment. However, there is not a ubiquitous radio signal propagation model due to complex phenomena (e.g., multi-path fading, shadowing, etc.) in indoor environment. Moreover, the performance of geometric-based scheme is vulnerable to be influenced by many factors, such as layout changes or people walking. On the contrary, fingerprinting-based scheme is more robust since it does not depend on radio signal propagation model. Typically, fingerprinting-based scheme consists of two phases: (1) construct fingerprint map, which firstly divides indoor space into a few cells and manually associates each cell with the scanned RSS values from surrounding APs; (2) online localization, which estimates the unknown location by matching the scanned RSS values with the fingerprinting map. The main bottleneck of fingerprint-based scheme is that manually constructing fingerprint map is time-consuming and labor intensive. For instance, the deployment overhead for a $300m^2$ environment is more than 7 hours [4]. Additionally, the fingerprinting map needs to be updated dynamically for maintaining localization accuracy.

For a practical subarea localization system, we argue several requirements are necessary: reasonable localization accuracy; no additional hardware components on user's side; scalable to large-scale deployment. On these basis, we propose a graph-based indoor subarea localization method with zero-configuration, which is infrastructure-free and constructing fingerprint map by passive crowdsourcing. Specifically, we firstly generate logical floor graph by utilizing two inherent characteristics of WiFi RSS in indoor environment, and then we formulate the problem of constructing fingerprinting map as a graph mapping problem between logical floor graph and physical floor graph. Finally, we utilize a Bayesian-based approach to estimate the unknown location.

978-1-5090-2771-2/16 $31.00 © 2016 IEEE
DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.41

The rest of this paper is structured as follows. Section 2 surveys related studies on indoor subarea localization. Section 3 describes our proposed method in detail. Section 4 reports and discusses our experimental results. Finally, we present our conclusion and future work in Section 5.

II. Related Work

In this section, we survey previous related works about indoor subarea localization and discuss how these works differ from our work. In general, existing studies on this topic can be divided into two categories:

A. Infrastructure-based Localization System

Infrastructure-based localization systems estimate unknown location based on the information from additional infrastructure or external equipment, such as WiFi signals, Bluetooth signals and ZigBee signals. For instance, the beacon frames from multiple Bluetooth APs [15] are used to localization the room, ZigBee interface [6] is used to collect WiFi RSS for room localization, wearable wrist sensors [7] is used to detect a person. The main drawback of infrastructure-based system is lack of scalability since costly infrastructure pre-deployment is necessary. Moreover, the performance of infrastructure-based systems is limited by disturbances and errors caused by indoor obstacles (e.g. walls, ceiling and furniture, etc.). Another challenge of infrastructure-based systems is how to design optimal configurations with trade-off the deployment cost and localization performance. [16] analyzed the localization performance and deployment issues by revealing localization error trends with geometric configurations, and concluded the optimal configuration is regular polygon where the vertices represent the RSS APs.

B. Infrastructure-free Localization System

In contrast, infrastructure-free localization systems utilize existing infrastructure (e.g., WiFi [10], [11], [17]–[19], magnetic field [20], etc.) to estimate an unknown location without deploying additional hardware.

Typically, localization methods for infrastructure-free system consist of geometric-based method and fingerprint-based method. Geometric-based method utilizes triangulation principle to estimate the unknown location based on radio propagation model, such as TOA [12], TDOA [13] and AOA [14]. However, there is not a ubiquitous radio propagation model in indoor environment, since the radio signal propagation would be strongly affected by multipath effect. In addition, specific devices for measuring TOA or AOA are costly. Fingerprint-based method utilizes the RSS values collected from a specific location as its fingerprint for labelling location. The localization process of this scheme includes two phases: construct fingerprint map and online localization. For example, [21] utilized fingerprint-based method with WiFi RSS to obtain room-level localization for visualizing indoor energy consumption. [22] proposed an subarea detection method using WiFi RSS. [23] proposed a more robust location fingerprint for localization using the RSS relative ordering of each pair of APs. For reducing erroneous estimation, [9] utilized the RSS characteristics when passing through a boundary point to calibration. However, previous fingerprinting-based method is infeasible because constructing fingerprint map is time-consuming and labor intensive [4].

Recently, some studies have been proposed to automatically construct fingerprinting map without time-consuming site survey. For instance, [18] proposed an indoor floor plan construction method with leveraging WiFi RSS and user motion information, which can be utilized to automatically construct fingerprinting map. WILL [10] automatically construct fingerprint map by utilizing RSS characteristics and user motions to . WicLoc [19] records user motions as well as WiFi signals for constructing fingerprint map. However, these methods need user's active participation when constructing fingerprint map. In contrast, our proposed method only utilize WiFi RSS to automatically construct fingerprint map, which can be done by passive crowdsourcing.

TABLE I: Notations used in indoor subarea localization

Symbol	Description
N, K, M	the num of subareas, WiFi APs, RSS traces
S, D, H	the set of subareas, WiFi RSS traces, Histogram bins
r^i, R	the RSS value from ap_i,the RSS values from all WiFi APs
$o(u, t, R)$	the RSS record collected by user u at time t
$L_i, traj(L_i)$	a WiFi RSS trace, a virtual trajectory
s_i, f_{si}	an indoor subarea,the fingerprint of subarea
ν_i	virtual subarea with high similarity fingerprint
Y	the fingerprint map
G_p, G_f	the physical floor graph, logical floor graph
τ	time windows size for identifying boundary points
σ	user-specific threshold for removing false identification

III. Graph-based Localization Method

In this section, we first introduce the key data structures and notations used in our proposed subarea localization method, and then present the problem definition and solution.

A. Problem Definition

For ease of the following presentation, we define the key notations used in the proposed method. Table I lists the relevant notations used in this paper.

Definition 1: RSS Record. A RSS record is a triple $o(u, t, R)$ that means the collected WiFi RSS values by user u at time t. R is a K dimensional vector and denote by $(r^1, ..., r^i, ..., r^K)$, r^i means the scanned WiFi RSS value from AP ap_i, K is the num of WiFi APs in indoor space and $1 \leqslant i \leqslant K$.

Definition 2: WiFi RSS Trace. We define a WiFI RSS trace as a set of RSS records and denote by $L = \{o_1, ..., o_i, ..., o_T\}$, o_i represents the collected RSS record at time t_i, $1 \leqslant i \leqslant T$.

Definition 3: Indoor Subarea. $S = \{s_1, s_2, ..., s_N\}$ denotes the set of subareas, N is the num of subareas and a subarea s_i refers to a region that makes up part of indoor space. Typically, subareas are rectangle, such as rooms and corridors, but not necessary.

Definition 4: Subarea Fingerprint. The feature of subarea s_i is defined as a $H \times K$ matrix $f_{si} = \{p_1, p_2, ..., p_K\}$, H is the histogram bins and p_j represents the histogram of scanned RSS values from ap_j in s_i, $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant K$.

We split the RSS values range into H bins and then p_j denote by a H dimensional vector, a bin-based method is used to calculate the p_j of subarea s_i, as shown in Equation. 1.

$$p_j = \prod_{h=1}^{H} \frac{\sum_{j=1}^{K} C_{ij}^h}{C_i} \tag{1}$$

Where $\sum_{j=1}^{K} C_{ij}^h$ is the num of collected RSS values from ap_j belongs to the h-th bin in total collected RSS values, C_i means the total collected RSS values in subarea s_i.

Definition 5: Fingerprint Similarity. The fingerprint similarity of subarea s_i and s_j is calculated by cosine similarity, as shown in Equation. 2.

$$Sim(f_{si}, f_{sj}) = \frac{1}{K} \sum_{n=1}^{K} \frac{Row_n(f_{si}) \cdot Row_n(f_{sj})}{||Row_n(f_{si})|| \times ||Row_n(f_{sj})||} \tag{2}$$

Where $Row_n(f_{si})$ and $Row_n(f_{sj})$ represent the n-th row vector of f_{si} and f_{sj}, respectively.

Definition 6: Fingerprint Map. The fingerprint map is a set of tuples by associating physical subarea and its fingerprint and denote by $Y = \{(s_1, f_{s1}, ..., (s_i, f_{si}), ..., (s_N, f_{sN}))\}$.

Definition 7: Physical Floor Graph. We denote the physical floor graph by $G_p =< V_p, E_p >$, where $V_p = \{v_1, v_2, ..., v_N\}$ and v_i represents subarea s_i, $E_p \subseteq V \times V$ correspond to the directly reachable of subareas in indoor space.

Based on the above definitions, we formulate the problem of indoor subarea localization as: Given: 1) indoor subarea set $S = \{s_1, s_2, ..., s_N\}$. 2) WiFi RSS Trace set $D = \{L_1, L_2, ..., L_M\}$ collected by passive crowdsourcing. 3) physical floor graph $G_p =< V_p, E_p >$. 4) a user localization request $o_?(u, t, R)$; Objective: find the correspond subarea s_i when scanning RSS record $o_?(u, t, R)$.

Our solution for this problem consists of two phases: (1) construct fingerprint map by graph mapping; (2) estimate the unknown subarea with a Bayesian approach.

B. Construct Fingerprint Map

In this subsection, we first give an high-level overview of our graph-based method for constructing fingerprint map, and then present the details of the method.

Unlike existing fingerprint-based methods, our method automatically constructs fingerprint map without manual site survey. First, we collect RSS traces by crowdsourcing (e.g., when participants go shopping, drink a coffee or relaxing). Then, after obtaining enormous RSS traces, the fingerprint map is constructed by the following three steps: modeling physical floor plan to an undirected graph, generate logical floor graph, and mapping logical floor graph to physical floor graph.

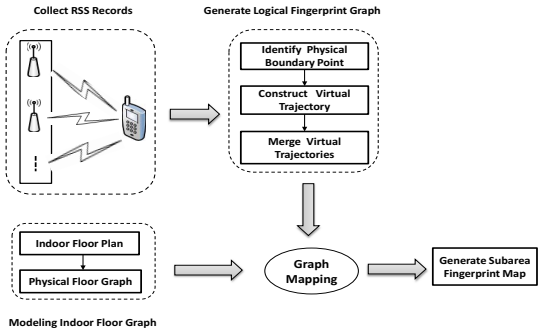


Fig. 1: High-level overview of constructing fingerprint map

1) Modeling Physical Floor Plan: Motivated by indoor robots pursuit/evasion research [24], we model the indoor floor plan with a undirected graph $G_p =< V_p, E_p >$ by decomposing the indoor floor plan into a collection of convex subareas, and further reduce the indoor space to a graph by discretization. Specifically, the discretization includes two steps:

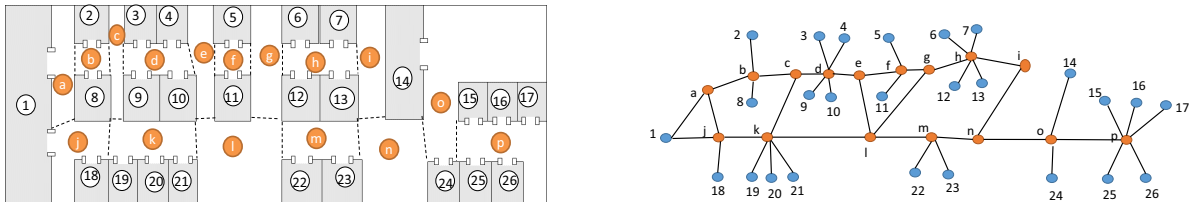
- *Step1*: decomposing the indoor floor plan into a set of convex subareas based on critical visibility events and association vertex v_i to subarea s_i;
- *Step2*: adding edges between vertices which are directly connected in the original indoor floor plan.

For example, the indoor floor plan of our experimental environment is shown in 2a, which consisting of 27 rooms and covering over $2000m^2$. Then, we decompose the floor plan into a set of subareas and add edges between directly connected vertices, and finally model the indoor floor plan as a undirected graph as shown in Figure 2b.

2) Generate Logical Floor Graph: A few factors can influence the propagation of radio signal in indoor environment, such as multiple diffraction, reflection of scattered signals from adjacent walls and crowd walking. By investigating spatial-temporal characteristics of indoor radio signal propagation, we observe two valuable characteristics can be exploited to subarea localization.

The first observation is physical obstacles, such as walls and stairs, will make WiFi RSS values jump dramatically. In order to investigate the physical obstacles effect on radio signal propagation, we collected 200 RSS records from three APs in room 1 and room 2, where AP1 and AP2 are located in room 1 and AP3 is located in room 2. The statistical information of RSS values is shown in Table II, and we can observe that the range of RSS values from the same AP significantly differ in different rooms.

Therefore, this characteristic can reflect the indoor floor plan to a certain degree and can be used to distinguish two subareas, which is also demonstrated in [25]. Based on this characteristic, we design a robust subarea fingerprint using RSS histogram as shown in *Definition 4*. In order to distinguish different subareas, we further define the similarity of Subarea Fingerprint as shown in Definition 5.



(a) indoor floor plan

(b) physical floor graph

Fig. 2: Modeling physical floor plan as a undirected graph

TABLE II: The RSS values scanned from three WiFi APs at different rooms

Range	AP1 at Room 1	AP1 at Room 2	AP2 at Room 1	AP2 at Room 2	AP3 at Room 1	AP3 at Room 2
$[-55,-40]$	115	0	93	1	0	120
$[-70,-55)$	72	3	81	5	4	63
$[-85,-70)$	10	11	21	17	21	15
$[-100,-85)$	3	23	5	39	42	2

Take RSS values of Table II as an example, split the range of RSS values into 4 bins: $\{(-40,-55],(-55,-70],(-70,-85],(-85,-100]\}$, the fingerprint of room 1 and room 2 can be calculated as f_{s1} and f_{s2}, respectively.

$$f_{s1} = \begin{bmatrix} 0.575 & 0.36 & 0.05 & 0.015 \\ 0.465 & 0.405 & 0.105 & 0.025 \\ 0 & 0.0597 & 0.3134 & 0.6269 \end{bmatrix} \quad (3)$$

$$f_{s2} = \begin{bmatrix} 0 & 0.0811 & 0.2973 & 0.6216 \\ 0.0161 & 0.0806 & 0.2742 & 0.629 \\ 0.6 & 0.315 & 0.075 & 0.01 \end{bmatrix} \quad (4)$$

The second observation is the WiFi RSS values will jump dramatically when passing a physical boundary point, such as room entrances and corners. For example, we collect a sequence of RSS values from three APs when walking from room 1 to room 2, as shown in Figure. 3a. Specifically, $\{t1,t2,t3,t4,t5\}$ are collected in room 1, $\{t6,t7,t8\}$ are collected when passing the entrance, $\{t9,t10,t11,t12\}$ are collected in room 2, as shown in Figure. 3b. We find that the "jump" range can reach 15dBm-30dBm. However, the RSS values should change smoothly in a small continuous area according to indoor empirical propagation model [26]. Therefore, the RSS "jump" characteristic when passing boundary points can be utilized to identify subarea entrance.

Based on the two spatial-temporal characteristics of radio signal propagation in indoor environment, we generate logical floor graph by three stages, as shown in Figure 4. Specifically, we first identify all physical boundary points based on the RSS "jump" characteristic when passing a physical boundary point, and remove false identification using subarea fingerprint similarity. Then, we partition a WiFi RSS trace into a virtual trajectory according to physical boundary points, as shown in Figure. 4. Finally, we merge all virtual trajectories to generate logical floor graph, as shown in Figure. 5.

Identify Physical Boundary Points. Based on the observation that the WiFi RSS values jump significantly when walking through a physical boundary point, we utilize the fluctuation of RSS values in a small time window to identify physical boundary points. Formally, given a WiFi RSS trace $L =< o_1,...,o_i,...,o_T >$, we define $Var(t_i,\tau)$ to represent the RSS fluctuation in time window $(t_i-\tau/2, t_i+\tau/2)$, as shown in Equation. 5.

$$Var(t_i,\tau) = \frac{1}{K}\sum_{i=1}^{K} Var(ap_i) \quad (5)$$

Where K is the number of WiFi APs, $Var(ap_i)$ is the variation of RSS values from ap_i during the time window, as calculated in Equation. 6.

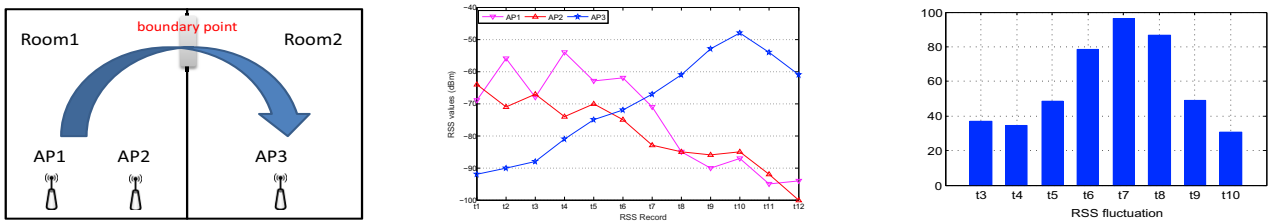
$$Var(ap_i) = \frac{1}{\tau-1}\sum_{j=t_i-\tau/2}^{t_i+\tau/2}(r_j^i - \overline{r^i})^2 \quad (6)$$

Where $\overline{r^i}$ is the average RSS values from ap_i in time window $(t_i-\tau/2, t_i+\tau/2)$, r_j^i is the RSS value from ap_i at time t_j.

If the RSS fluctuation in time window $(t_i-\tau/2, t_i+\tau/2)$ is significantly higher than average, we can infer the user is walking through a physical boundary point at time t_i. Formally, we use variation coefficient α to quantify the degree of RSS "jump", as shown in Equation. 7.

$$\alpha = \frac{\tau \times Var(t_i,\tau)}{\sum_{j=t_i-\tau/2}^{t_i+\tau/2} Var(t_j,\tau)} \quad (7)$$

For example, set time window size $\tau = 5$ and variation coefficient as $\alpha = 1.3$, the variation of RSS values from three APs in Figure 3b is calculated as shown in Table III. We further calculate the RSS fluctuation: $V = \{36.97, 34.6, 48.4, 78.67, 96.4, 86.8, 49.27, 30.77\}$ as shown



(a) (b) (c)

Fig. 3: The "jump" characteristic when passing physical boundary points

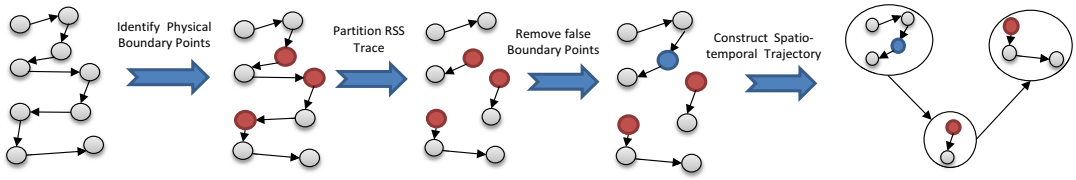


Fig. 4: Construct virtual trajectory of WiFi RSS trace

TABLE III: The variation of RSS values from three WiFi APs

time window	$(t1, t5)$	$(t2, t6)$	$(t3, t7)$	$(t4, t8)$	$(t5, t9)$	$(t6, t10)$	$(t7, t11)$	$(t8, t12)$
AP1	46.5	31.8	42.3	137.5	162.7	143.5	80.8	18.7
AP2	14.7	10.3	36.7	40.3	48.7	20.2	11.7	42.3
AP3	49.7	61.7	66.3	58.2	77.8	96.7	55.3	31.3

in Figure 3c, and infer the user is passing a physical boundary point in time $\{t_6, t_7, t_8\}$.

Remove False Identification. As mentioned above, we identify physical boundary points according to the RSS "jump" characteristic. However, this method may bring some false positives, since other factors (e.g., crowd passing and furniture layout change, etc.) may create similar RSS "jump". However, subarea fingerprint using RSS histogram is stable and robust according to the first observation. On the basis, we remove false positives based on the similarity of subarea fingerprint.

Formally, after obtaining time set $\Omega = \{t_p, t_{p+1}, ..., t_q\}$ that users may walk through physical boundary points according to RSS "jump" characteristic, we partition RSS trace L into a subsequence set $L = \{o(t_1 : t_p), o(t_p : t_{p+1}), ..., o(t_{q-1} : t_q), o(t_q : t_T)\}$, $o(t_p : t_{p+1})$ is the RSS subsequence collected from t_p to t_{p+1}. Then, we calculate the fingerprint of each RSS subsequence as denote by $F = \{f_p, f_{p+1}, ..., f_q\}$, f_p represents the fingerprint of RSS subsequence $o(t_1 : t_p)$. Finally, we use a threshold-based approach to remove false positives, which means t_{p+1} is a false positive if the fingerprint similarity between f_p and f_{p+1} is greater than a threshold δ, as shown in Equation. 8.

$$Sim(f_p, f_{p+1}) > \delta \tag{8}$$

Construct Virtual Trajectory. After removing false identification of physical boundary points, we repartition the RSS trace L into a subsequence set $L = \{o(t_1 : t_p), o(t_p : t_{p+1}), ...\}$ and map each RSS subsequence $o(t_p : t_{p+1})$ to a virtual subarea ν_{p+1}. A virtual subarea is a container which consists of fingerprint with high similarity. Finally, we construct the virtual trajectory of RSS trace L as $traj(L) =< \nu_p \rightarrow \nu_{p+1} \rightarrow ... >$, as shown in Figure. 4.

Generate Logical Floor Graph. After constructing virtual trajectory for each RSS trace, we generate logical floor graph $G_f(V_f, E_f)$ by merging all virtual trajectories $\{traj(L_1), traj(L_2), ..., traj(L_M)\}$. Specifically, the merge process consists of two steps:

- *Step1:* using K-means to cluster virtual trajectories $\{traj(L_1), traj(L_2), ..., traj(L_M)\}$ into P clusters, and mapping cluster center π_i of cluster P_i to vertex v_i of logical floor graph, as shown in Figure. 5b. In the clustering process, using fingerprint similarity (See in Definition 5) to measure the closeness of two virtual subareas.
- *Step2:* adding an edge between v_i and v_j if cluster P_i and cluster P_j is reachable, which means that there is at least one pair of adjacent virtual subareas $< \nu_i \rightarrow \nu_j >$ for $\forall \nu_i \in P_i$ and $\forall \nu_j \in P_j$, as shown in Figure. 5c.

3) Mapping Logical Floor Graph to Physical Floor Graph: For automatically constructing fingerprint map, we need to associate virtual subarea ν_i to the corresponding subarea s_j by mapping logical floor graph to physical floor graph. Formally, given logical floor graph $G_f =< V_f, E_f >$ and physical floor graph $G_p =< V_p, E_p >$, find a mapping function $\tau : V_f \rightarrow V_p$ for $\forall e(u, v) \in E_f, e(\tau(u), \tau(v)) \in E_p$. Obviously, this is a subgraph isomorphism problem and can be solved by Ullmann algorithm [27].

Ullmann algorithm utilizes a depth-first search strategy to

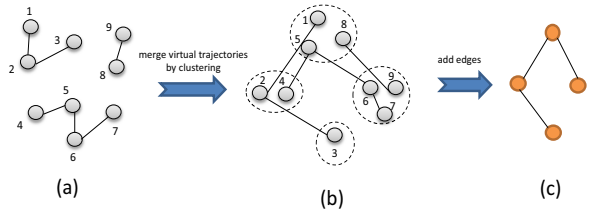


Fig. 5: Construct logical floor graph

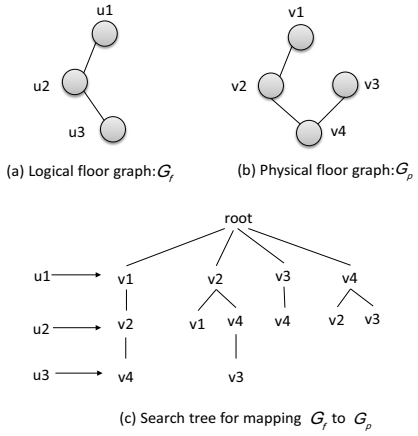


Fig. 6: Mapping logical floor graph to physical floor graph

enumerate all sub-graphs of G_f that matching G_p. For ease of understanding, Figure 6c is the search tree for mapping G_f (Figure 6a) to G_p (Figure 6b), the i-th layer of search tree represents mapping u_i of G_f to each node of G_p, a path from root node to leaf node represents a subgraph matching between G_p and G_f. A subgraph matching is correct if the adjacency relationship of u_i in G_f is the same as its mapping node v_j in G_p.

Since we have mapped each virtual subarea ν_i to the corresponding physical subarea s_j, we further compute the fingerprint of s_j according to Equation. 1. Then, we construct subarea fingerprint map with associating s_j to the calculated fingerprint.

C. Online Localization

At the online localization part, user sends localization request with submitting the scanned RSS record $o(u,t,R)$, $R = \{r^1, r^2, ..., r^K\}$, our method estimates the subarea of his/her current location using a Bayesian approach. According to Bayesian inference, the posterior probability $P(s_i|R)$ can be calculated as Equation. 9.

$$P(s_i|R) = \frac{P(R|s_i)P(s_i)}{P(R)} \tag{9}$$

Since the prior probability that user is located in each subarea is equal and the RSS values from different WiFi APs are independent, the posterior probability $P(s_i|R)$ can further be simplified as Equation. 10.

$$P(s_i|R) \propto \prod_{j=1}^{K} P(r^j|s_i) \tag{10}$$

For a given subarea s_i, the prior probability $P(r^j|s_i)$ can be calculated by the normalized histogram of ap_j in this subarea. We partioned the RSS values range into H bins when constructing fingerprint map, suppose r^j belongs to the h-th bin, $P(r^j|s_i)$ is equal to $f_{si}(h,j)$. Then, the localization result for RSS record $o(u,t,R)$, $R = \{r^1, r^2, ..., r^K\}$ can be estimated by Equation. 11.

$$\widehat{s} = \underset{s_i \in S}{argMax} \prod_{j=1}^{K} f_{si}(h,j) \tag{11}$$

Algorithm 1 formally describes the framework of our proposed method for indoor subarea localization. First, as shown in Lines $2 \sim 7$, we generate the logical floor graph based on two unexploited RSS characteristics in indoor space. Then, as depicted in Line $8 \sim 9$, we construct subarea fingerprint map by mapping logical floor graph to physical floor graph. At the online localization part, we calculate the posterior probability for each subarea based by Bayesian inference, as shown in Line $11 \sim 14$. Finally, we select the subarea with the maximum posterior probability as the localization result.

Algorithm 1 Graph-based method for indoor subarea localization

Require: 1) The RSS traces set $D = \{L_1, L_2, ..., L_M\}$; 2) Subarea set $S = \{s_1, s_2, ..., s_N\}$; 3) Physical floor graph G_p; 5) user-specific threshold: τ, α, δ; 4)The RSS record of user's localization request: $o < u, t, R >$ and $R = \{r^1, r^2, ..., r^K\}$.

Ensure: The subarea s_u of user's current location

1: $***$Phase 1: Construct Fingerprint Map$***$
2: **for** $\forall L_i \in D$ **do**
3: Identify physical boundary points according to Equation. 7.
4: Remove false identification according to Equation. 8.
5: Construct virtual trajectory $traj(L_i)$.
6: **end for**
7: Generate logical floor graph G_f by merging virtual trajectories $\{traj(L_1), traj(L_2), ..., traj(L_M)\}$.
8: Map logical floor graph G_f to physical floor graph G_p.
9: Construct subarea fingerprint map $Y = \{(s_1, f_{s1}, ..., (s_i, f_{si}), ..., (s_N, f_{sN}))\}$.
10: $***$Phase 2: online localization$***$
11: **for** $\forall (s_i, f_{si}) \in Y$ **do**
12: Otain the histogram bin h that r^j belongs to.
13: Calculate the probability $P(s_i|R) = \prod_{j=1}^{K} f_{si}(h,j)$
14: **end for**
15: **return** s_u=arg $\underset{s_i \in S}{Max} P(s_i|R)$.

TABLE IV: The RSS sample format

001-123	124	125	126	127
RSS values	timestamp	phone ID	boundary point flag	subarea ID

TABLE V: One example of RSS sample

[001]	...	[123]	[124]	[125]	[126]	[127]
-73	...	-87	2015-12-07 15:28:15	1	0	1

IV. Experiment Evaluation

In this section, we first describe the experimental setting and dataset for evaluation. Then, we report the results of a series of experiments conducted to evaluate the performance of our proposed method for indoor subarea localization, follow by discussions.

A. Experimental Setup

Our experimental environment is a large indoor shopping mall with 26 shops and 7 corridors. Each shop is regarded as a subarea and corridors are partitioned to 16 subareas, so there are 42 subareas in total. The floor plan and subarea partition are shown in Figure 2. To evaluate our subarea localization method, we need to record two labeled information: the subarea and whether the location is a physical boundary point of each WiFi RSS record. We develop a mobile application to collect WiFi RSS samples with a sampling rate of 1 Hz, each sample is represented by a tuple: $< L, o >$. Specifically, $L = \{s_i, 0|1\}$ is the label information of subarea and whether is a physical boundary point, o is the scanned RSS record from surround WiFi APs and represented by a triple $(M, t, < r^1, r^2, ..., r^K >)$, M is the MAC address of collection device and t is the collection time, r^1 is the scanned RSS values from AP1. Note that we collect RSS information with a sampling rate of 1 Hz at the offline phase for constructing fingerprint map, users only need to submit the single RSS sample in online localization without continuously collecting RSS information.

B. Experimental Datasets

We collect 117 WiFi RSS traces for experiment evaluation by 25 participants (including students and shop workers) over 33 days, in which one RSS trace includes an average of 10 subareas and 1532 RSS records, and each subarea has been visited by at least three participants. Statistically, there are 123 different WiFi APs and 179241 WiFi records in this dataset. For constructing subarea fingerprint and calculating fingerprint similarity, we extent each RSS sample to a 128 dimensional vector, as shown in Table IV. For WiFi AP without collecting RSS values, we set -110 dBm as default value, one example of RSS samples is shown in Table V.

C. Experimental results

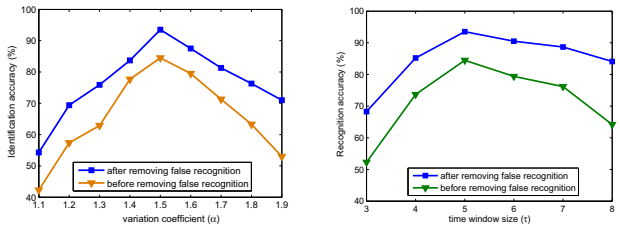
1) Identify Physical Boundary Points: Three parameters in our algorithm need to be determined for identifying physical boundary points: time windows size τ, variation coefficient α for recognition boundary points, user-specific threshold δ for removing false identification. The three parameters directly impact the accuracy of identifying physical boundary Points. We use a cluster-based method to select δ. Specifically, we first cluster all WiFi RSS records to N classes by KNN, N is the num of subareas. Then, we calculate fingerprint of each class and further obtain the similarity for each pair of fingerprints. Finally, we select the average similarity as δ for removing false identification.

TABLE VI: The accuracy of identifying physical boundary points with different time window sizes and variation coefficients

τ \ α	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
3	0.23	0.37	0.44	0.47	0.52	0.37	0.29	0.24	0.19
4	0.29	0.48	0.51	0.59	0.71	0.64	0.48	0.41	0.21
5	0.43	0.56	0.61	0.75	0.83	0.76	0.70	0.63	0.53
6	0.37	0.45	0.51	0.70	0.79	0.57	0.46	0.47	0.33
7	0.30	0.37	0.44	0.59	0.67	0.55	0.39	0.35	0.20
8	0.24	0.29	0.32	0.48	0.54	0.38	0.31	0.21	0.16

For calculating the subarea fingerprint, we partition the range of RSS values into 4 bins which is in line with typical RSS quality partition [28], [29]: (1) bin-1, which represents WiFi signal is excellent and the RSS values are in range [-55,0]; (2) bin-2, which represents WiFi signal is good and the RSS values are in range [-70,-55); (3) bin-3, which represents WiFi signal is poor and the RSS values are in range [-85,-70); (4) bin-4, which represents WiFi signal is bad and the RSS values are in range [-100,-85).

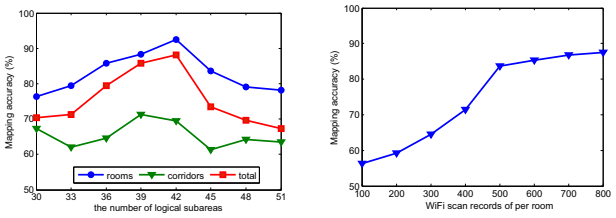
Table VI shows the accuracy of identifying physical boundary points with time window size τ and variation coefficient α. From this table, we observe: 1) the accuracy drops sharply when the user-specific threshold of variation coefficient α is lower than 1.2 or greater than 1.5; 2) Set $\alpha = 1.3$, the accuracy increases with time window size increasing from 1 to 5, and slightly decrease when the time window size is larger than 5 due to the RSS fluctuation between physical boundary point and other location will be smaller for a large time window size. Finally, the best performance (83%) is achieved when $\alpha = 1.3$ and $\tau = 5$.



(a) The accuracy with different variation coefficients

(b) The accuracy with different time window sizes

Fig. 7: Parameter tuning for identifying physical boundary points



(a) The mapping accuracy with different virtual subareas

(b) The mapping accuracy with different RSS records of per subarea

Fig. 8: Parameter tuning for mapping accuracy

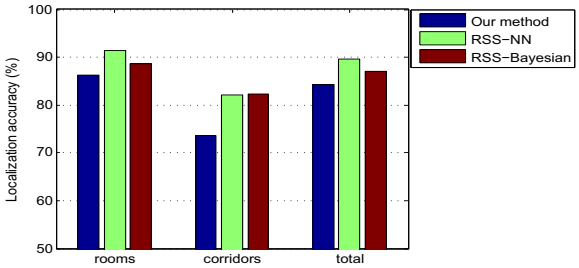


Fig. 9: The subarea localization accuracy

Figure 7a and Figure 7b show the identification accuracy as a function of variation coefficient and time window size, respectively. From the two figures, we observe: 1) the method using subarea fingerprint similarity can effectively remove false recognition; 2) Set the time window size $\tau = 5$, the accuracy declines sharply when variation coefficient α is greater than 1.6 or lower than 1.4, and achieve the best accuracy when $\alpha = 1.5$; 3) Set $\alpha = 1.5$, the identification accuracy increases with the increasing number of time window size between 3 and 5, and slightly decrease when the time window size is larger than 5; 4) the performance of removing false identification decreases slightly with increasing time window size, due to the difference of RSS fluctuation between physical boundary point and normal location will be smaller with increasing time window size.

2) Construct Fingerprint Map: We utilize mapping accuracy to evaluate the performance for constructing fingerprint map. The mapping accuracy (MA) is defined in Equation 12. We define s_i as the ground truth subarea label of record o_i, $\widehat{s_i}$ is the mapping subarea label.

$$MA = \frac{\sum_{i=1}^{Te} I(si, \widehat{s_i})}{Te} \tag{12}$$

Where $I(si, \widehat{s_i})$ is an indicator function that return 1 if $\widehat{s_i} = s_i$, Te is the test RSS records for evaluation.

Figure 8a reports the performance of constructing fingerprint map with different parameter settings. One parameter need to be determined for constructing fingerprint map: the cluster number K_f for generating logical floor graph. As shown in Figure 8a, we show the performance where K_f is in the range [30,33,...51]. From this figure, we can see that the mapping accuracy increases gradually when K_f increases from 30 to 42 and then drops when K_f is greater than 42, the highest mapping accuracy is 92.3% when K_f equals to 42 (the number of physical subareas). Another observation is the mapping accuracy for subareas located in corridor is lower about 20 percent than rooms, which shows there no obvious RSS "jump" characteristic for two connected subareas in corridor because there are no walls or physical boundary points can significantly weakened the radio signal strength.

Figure 8b reports the performance of constructing fingerprint map as a function of number of WiFi RSS records per subarea. We can see that the mapping accuracy is relatively stable when RSS records of each subarea is more than 500, which shows our algorithm for constructing the fingerprint map will converge quickly and has a low crowdsourcing data requirement. Moreover, the performance of constructing fingerprint map will improve with increasing collected data.

3) Localization Accuracy: We evaluate the performance of the proposed localization method by comparing with two well-known subarea localization methods. We first introduce the experimental dataset and parameters setting, then detail the comparative localization techniques. Finally, we report and discuss the experimental results.

Dataset. We randomly select 70% RSS records of each subarea as training dataset to construct fingerprint map, and the rest 30% as testing dataset for evaluation localization accuracy.

Parameters Setting. Tuning algorithm parameters, such as the time window size for identification physical boundary points and the clusters for constructing logical floor map, are critical to the performance of localization. According to the experience of previous experiments, our algorithm empirically set parameters as: $\{\tau = 5, \alpha = 1.5, K_f = 42\}$, for constructing fingerprint map.

Comparative Methods. We compare our method with the following two methods that have been widely used in subarea localization: (1) RSS-NN [17], which constructs fingerprint map by manual site survey and estimates subarea using KNN classification; (2) RSS-Bayesian [9], which also constructs fingerprint map by site survey and estimates subarea using Bayesian inference.

Results and Analysis. Figure 9 shows the localization accuracy of the three methods. It can be seen that the performance for open subarea (subareas in the corridor) and closed subarea (room) are significantly different for all methods. As shown in Figure 9, the localization accuracy of rooms are more than 87% for the three methods, but lower than 85% for open subareas in corridor, which shows RSS values of two connected open subareas are too similar to distinguish. RSS-NN achieves the best performance for both closed subarea(92%) and open subarea(83%). Another observation is the average localization accuracy rate is 85% for our method, which is 4% less than RSS-NN. Therefore, our method can obtain considerable performance compared with previous methods with labor intensive and time-consuming site survey.

V. Conclusion

Indoor subarea localization has attracted a few research efforts from both academia and industry in recent years . This paper has proposed a ready-to-deploy method for indoor subarea localization with zero-configuration, since the proposed method is infrastructure-free and does not need time-consuming site survey. The main idea is to generate logical floor graph based on two characteristics of WiFi RSS in indoor space, and automatically construct fingerprint map by mapping logical floor graph to physical floor graph. The proposed method has been implemented and deployed in a real-world shopping mall with 25 users over 33 days with an average localization accuracy of 85%, which is competitive to traditional approaches. The advantages on infrastructure-free and automatically constructing fingerprint map, make our method can be widely used in indoor environment.

Acknowledgment

This work is partially sponsored by the National Basic Research 973 Program of China (No. 2015CB352403), the National Natural Science Foundation of China (NSFC) (61261160502, 61272099), the Program for National Natural Science Foundation of China / Research Grants Council (NSFC/RGC)(612191030), the Program for Changjiang Scholars and Innovative Research Team in University (IRT1158, PCSIRT), the Scientific In-novation Act of STCSM (13511504200), and EU FP7 CLIMBER project (PIRSES-GA-2012-318939).

References

[1] J. She, J. Crowcroft, H. Fu, and F. Li, "Convergence of interactive displays with smart mobile devices for effective advertising: A survey," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 10, no. 2, p. 17, 2014.

[2] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 579–590, 2013.

[3] H. Shin, Y. Chon, Y. Kim, and H. Cha, "A participatory service platform for indoor location-based services," *Pervasive Computing, IEEE*, vol. 14, no. 1, pp. 62–69, 2015.

[4] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, S. Sen, and V. Handziski, "Microsoft indoor localization competition: Experiences and lessons learned," *GetMobile: Mobile Computing and Communications*, vol. 18, no. 4, pp. 24–31, 2015.

[5] E. Leitinger, M. Fröhle, P. L. Meissner, and K. Witrisal, "Multipath-assisted maximum-likelihood indoor positioning using uwb signals," in *Communications Workshops (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 170–175.

[6] Y. Gao, J. Niu, R. Zhou, and G. Xing, "Zifind: Exploiting cross-technology interference signatures for energy-efficient indoor localization," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2940–2948.

[7] J. Ranjan and K. Whitehouse, "Object hallmarks: Identifying object users using wearable wrist sensors," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 51–61.

[8] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 293–304.

[9] S. Hotta, Y. Hada, and Y. Yaginuma, "A robust room-level localization method based on transition probability for indoor environments," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012, pp. 1–8.

[10] C. Wu, Z. Yang, Y. Liu, and W. Xi, "Will: Wireless indoor localization without site survey," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 839–848, 2013.

[11] Z. Zheng, Y. Chen, T. He, L. Sun, and D. Chen, "Feature learning for fingerprint-based positioning in indoor environment," *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.

[12] J. Xiong, K. Sundaresan, and K. Jamieson, "Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 537–549.

[13] S. Pang and R. Trujillo, "Indoor localization using ultrasonic time difference of arrival," in *Southeastcon, 2013 Proceedings of IEEE*. IEEE, 2013, pp. 1–6.

[14] Y. Hou, Y. Xue, C. Chen, and S. Xiao, "A rss/aoa based indoor positioning system with a single led lamp," in *Wireless Communications & Signal Processing (WCSP), 2015 International Conference on*. IEEE, 2015, pp. 1–4.

[15] P. Martin, B.-J. Ho, N. Grupen, S. Munoz, and M. Srivastava, "An ibeacon primer for indoor localization: demo abstract," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 190–191.

[16] A. Hossain and W.-S. Soh, "Cramer-rao bound analysis of localization using signal strength difference as location fingerprint," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[17] S. Zhou, B. Wang, Y. Mo, X. Deng, and L. T. Yang, "Indoor location search based on subarea fingerprinting and curve fitting," in *High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), 2013 IEEE 10th International Conference on*. IEEE, 2013, pp. 2258–2262.

[18] Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan, "Hallway based automatic indoor floorplan construction using room fingerprints," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 315–324.

[19] J. Niu, B. Wang, L. Cheng, and J. J. Rodrigues, "Wicloc: An indoor localization system based on wifi fingerprints and crowdsourcing," in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3008–3013.

[20] M. Angermann, M. Frassl, M. Doniec, B. J. Julian, and P. Robertson, "Characterization of the indoor magnetic field for applications in localization and mapping," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012, pp. 1–9.

[21] N. Castelli, G. Stevens, T. Jakobi, and C. Ogonowski, "Placing information at home: using room context in domestic design," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 919–922.

[22] K. Hida, C. Bin, Y. Hada, and S. Mori, "Evaluation of area detection method using machine learning," in *Multimedia, Distributed, Cooperative, and Mobile Symposium*, 2014.

[23] J. T. Biehl, M. Cooper, G. Filby, and S. Kratz, "Loco: a ready-to-deploy framework for efficient room localization using wi-fi," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 183–187.

[24] A. Kehagias, G. Hollinger, and S. Singh, "A graph search algorithm for indoor pursuit/evasion," *Mathematical and Computer Modelling*, vol. 50, no. 9, pp. 1305–1317, 2009.

[25] R. De Francisco, "Indoor channel measurements and models at 2.4 ghz in a hospital," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. IEEE, 2010, pp. 1–6.

[26] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[27] J. R. Ullmann, "An algorithm for subgraph isomorphism," *Journal of the ACM (JACM)*, vol. 23, no. 1, pp. 31–42, 1976.

[28] A. S. Azini, M. R. Kamarudin, and M. Jusoh, "Transparent antenna for wifi application: Rssi and throughput performances at ism 2.4 ghz," *Telecommunication Systems*, pp. 1–9, 2015.

[29] E. Peltonen, E. Lagerspetz, P. Nurmi, and S. Tarkoma, "Energy modeling of system settings: A crowdsourced approach," in *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*. IEEE, 2015, pp. 37–45.