Fingerprinting Based Indoor Localization Considering the Dynamic Nature of Wi-Fi Signals

Nasim Alikhani, Vahideh Moghtadaiee, Seyed Ali Ghorashi

Abstract

Current localization techniques in outdoors cannot work well in indoors. Wi-Fi fingerprinting technique is an emerging localization technique for indoor environments. However in this technique, the dynamic nature of WiFi signals affects the accuracy of the measurements. In this paper, we use affinity propagation clustering method to decrease the computation complexity in location estimation. Then, we use the least variance of Received Signal Strength (RSS) measured among Access Points (APs) in each cluster. Also we assign lower weights to altering APs for each point in a cluster, to represent the level of similarity to Test Point (TP) by considering the dynamic nature of signals in indoor environments. A method for updating the radio map and improving the results is then proposed to decrease the cost of constructing the radio map. Simulation results show that the proposed method has 22.5% improvement in average in localization results, considering one altering AP in the layout, compared to the case when only RSS subset sampling is considered for localization because of altering APs.

Key words: fingerprinting; indoor localization; Received Signal Strength (RSS); altering Access Point (AP); clustering.

1. INTRODUCTION

Nowadays, Indoor location based services (LBSs) have attracted serious attention among researchers [1,2]. Indoor LBSs significantly improve and expand the network management and security [3,4], including personal navigation for increasing safety in emergency operations, healthcare monitoring on visiting patients, locating resources, for instance medical records, or consulting with other specialists, delivery in personal information [5], context awareness in providing personalized services for mobile users over the internet with applications on a meeting in another company or a business trip [3]. Generating a responsive model for big areas such as shopping malls, airports, universities needs a lot of cost in time and energy to overcome the details of these areas for accurate localization [4].

Global Positioning Systems (GPS) cannot operate well especially in indoor areas [6,7] because of reflection and scattering issues caused by obstacles [7]. The visibility of satellite signals in indoor environments is considerably lower than in outdoors. This challenge limits the accuracy of LBSs in indoor environments, where people spend nearly up to 89% of their time. This in turn affects the user experience, especially for the small screens of smart-phones [1]. Therefore, range based localization techniques are used that employ measurement techniques such as time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA) to estimate the distance between two nodes. This techniques give accurate information only when line-of-sight (LOS) connection is available [8]. The time-based raging techniques require synchronization and AOA needs high cost hardware for implementation [8]. Usually, some non-satellite-based signal sources such as Wi-Fi are preferred for indoor localization [9] that need only Access Points (APs) as low cost hardware infrastructures already available in the area. However, other signals such as Bluetooth, FM radio [10], radio-frequency identification (RFID) and magnetic field are also used in indoor localization.

Smartphones are everywhere and known as a promising infrastructure for all the existing and future services. Smartphones are also used for indoor location services where Wi-Fi is used as a localization source [11]. The requirement of Wi-Fi networks is widely spread in public and private buildings [1]. In this paper we use Wi-Fi as signal sources for indoor localization. Extracting range measurements using RSS is easy and cheap to implement and therefore, RSS- based location fingerprinting is very popular in indoor localization [1,12]. In RSS fingerprinting, normally, a set of measured RSS that are unique for location spots, serve as a fingerprint.

Wireless Local Area Network (WLAN) fingerprinting consists of two phases: offline and location estimation. In the offline phase, multiple RSS at each Reference Point (RP), from each AP are measured throughout a time interval and these fingerprint vectors are stored along with their location coordinates in a radio map. In location estimation phase, Test Point (TP) measures the RSS vector at specified location and applies pattern recognition algorithms to associate these measurements to the radio map by finding similar fingerprints [1,12]. This technique has some challenges that should be addressed by network designers. In [13] a survey on these challenges and some methods to deal with them are examined. For example, environmental variations such as altering AP signals with time (generally in weeks or months) [14,15] may decrease the accuracy of localization. Altering APs is the effect of changing AP signals and it caused by AP movement, wall partitioning and variations on AP signals by obstacles in indoors. Also diverse devices in both phases of fingerprinting method can reduce the accuracy of localization. Therefore, methods based on using cosine similarity and considering the least variance among APs between RSS of RPs and TP can reduce the localization error significantly [16,17], as various smartphones have differences on their hardware. However, constructing the radio map has a high cost in time and energy.

Using all measured RSS values at RPs increases the complexity of computation, therefore pre-processing techniques are employed to reduce the search space of the TP location into a smaller region and delete outlier data for higher accuracy in location estimation phase [13]. Therefore, we can arrange the features of RSS measured by clustering method. K-means clustering algorithm [13] is a popular method between various clustering techniques, however this method is sensitive to the initial selection of cluster heads (CHs). These CHs are chosen randomly at the first time. Therefore we must rerun this algorithm to find a good solution. One of the clustering method that has solved this problem is affinity propagation clustering algorithm [18].

Measured RSS by different smartphones has different value, which affects the accuracy of localization. Therefore, device diversity in both phases of fingerprinting technique is a challenging issue that needs to be addressed [15,16]. Authors in [17] proposed a method that estimates the cluster where TP belongs to, by a method that considers the least variance of RSS fingerprint vectors among APs as a candidate CH for TP in room level localization and can improve the results of localization by using diverse devices in the location estimation phase. However, this method did not consider altering APs and did not use radio map updating algorithm in order to improve the results of localization. Authors in [19] proposed an algorithm that use a weighting factor for some closer points to TP in terms of signal distance. Then, they use a method to delete the outlier data. Therefore in this method, despite considering the unequal trend of signal distance and physical distance in indoors, it did not consider altering APs and radio map updating.

The Simultaneous Localization and Mapping (SLAM)-based approach utilized inertial smartphone sensors in order to improve the signal map. These approaches have errors in measuring signal and they must be calibrated at any time with high cost [20,21]. Methods based on using Gaussian process latent variable models can relate the RSS fingerprints and simulate human movements (displacement, direction, etc.) as hidden variables [22]. However, these methods often consider a relatively stable radio map without altering APs. Authors in [23] learned the functional relationship between the initial radio map and real time readings by using nonlinear regression analysis and using the nearest neighbor method to find locations. However, this method assumed a certain trend in AP signals, therefore, these algorithms may not increase the accuracy by AP altering assumptions. Authors in [24] have applied a regression algorithm on RSS received at RPs in order to construct the radio map. This method had worked with the same APs in the radio map. However, it is difficult to estimate the accurate regression model when we have altering APs. Authors in [25] proposed a radio map generation method that could improve the results of localization by polishing the input RSS data in radio map by a semi-supervised algorithm, and they used a method for updating the radio map. However, they did not use any algorithm for improving the results of localization by considering the trend of altering AP signals. Authors in [15] used RSS subset sampling because of altering APs and their variations on time. However, this method did not consider the least variance of RSS measured among APs. This method did not consider a weight for each point in each cluster in signal space because of obstructions and alterations in indoors.

The proposed method in this paper can improve the indoor localization results by handling altering APs and signal changes. This method localizes a TP by considering the lowest variance of RSS value among APs, analyses the existence of obstacles in indoors by assigning a lower weight to altering AP that represents the signal distance is not similar to physical distance and combines with RSS subset sampling method that is mentioned in [15]. Our proposed method is compared with KNN [26] as a deterministic algorithm and Gaussian kernel algorithm [27] as a probabilistic algorithm. Also we compare this method with RSS subset sampling method [15]. Our proposed method also handles the device diversity problem and does not need any calibrated sensor device. Furthermore, we use an updating radio map algorithm in order to improve the results. The results of our measured RSS data are analyzed with and without considering altering AP. Also, the results of three simulated radio maps with various number of APs are analyzed.

The organization of the paper is as follows. Section 2 determines the fingerprinting system model, formulations and algorithms that are used in this paper. In section 3 the proposed method is explained. Section 4 examines the considered layout area and the simulation parameters. In section 5, we simulate all mentioned methods and our proposed method. Section 6 concludes the paper.

2. System Model and Formulation

In this section, we explain two phases of fingerprinting localization method. We express the algorithms used in this paper for comparison with our proposed method. Then we examine the path-loss formula for our simulations purposes.

2.1. Offline Phase

The first phase in fingerprinting localization method is offline phase. In this phase, we consider L RPs as $\{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_j, \dots, \mathbf{l}_L\}$ in two dimensions with known values such as $\mathbf{l}_j = (x_j, y_j)$. There are T measured RSS values from each AP by each RP. The RSS fingerprint vector for the j^{th} RP is $[s_{\mathbf{l}_j,AP_1}(t), s_{\mathbf{l}_j,AP_2}(t), \dots, s_{\mathbf{l}_j,AP_i}(t), \dots, s_{\mathbf{l}_j,AP_{AP}}(t)]$ that is stored along with his location in radio map. $s_{\mathbf{l}_j,AP_i}(t)$ is the measured RSS value from the i^{th} AP in the j^{th} RP in the t^{th} unit time. The variable for unit time is $t = 1,2, \dots, T$. The time averaged RSS vector at location $\mathbf{l}_j$ is defined by $\mathbf{S}_{\mathbf{l}_j} = [\bar{s}_{\mathbf{l}_j,AP_1}; \bar{s}_{\mathbf{l}_j,AP_2}; \dots; \bar{s}_{\mathbf{l}_j,AP_i}; \dots; \bar{s}_{\mathbf{l}_j,AP_{AP}}]$.

$$\bar{s}_{\mathbf{l}_j,AP_i} = \frac{1}{T}\sum_{t=1}^{T} s_{\mathbf{l}_j,AP_i}(t) \qquad (1)$$

Therefore, by these vectors, the RSS values in the radio map can be shown as (2).

$$\mathbf{S}_{\mathbf{l}} = \begin{pmatrix} \bar{s}_{\mathbf{l}_{1},AP_1} & \cdots & \bar{s}_{\mathbf{l}_{1},AP_n} \\ \vdots & \ddots & \vdots \\ \bar{s}_{\mathbf{l}_{L},AP_1} & \cdots & \bar{s}_{\mathbf{l}_{L},AP_n} \end{pmatrix} \tag{2}$$

2.2. Location Estimation Phase

The second phase of the fingerprinting localization method is location estimation phase. The unknown location of TP is $l = (x, y)$. For estimating the location of TP in the second phase of the fingerprinting method, first we measure the RSS vector from all APs at the specified location of TP. The objective is to estimate the location of TP that is denoted by $\hat{l}$. $\mathbf{S}_l' = [\bar{s}_{l,AP_1}; \bar{s}_{l,AP_2}; \dots; \bar{s}_{l,AP_i}; \dots; \bar{s}_{l,AP_{AP}}]$ is the RSS vector received by TP, then by pattern recognition techniques the location of TP can be estimated [13].

2.2.1. KNN and KWNN Algorithms

The KNN and KWNN methods are deterministic methods which can estimate the location of TP by calculating respectively the average and weighted average of the coordinates of *K* nearest points [26]. The weighting values in KWNN method are the inverse of the Euclidean distance that are multiplied by the coordinates of *K* nearest points. The Euclidean distance from RSS of TP and the j^{th} RP is shown in (3). If we assume $\mathbf{S}_l' = [\bar{s}_{l,AP_1}; \bar{s}_{l,AP_2}; \dots; \bar{s}_{l,AP_i}; \dots; \bar{s}_{l,AP_{AP}}]$ is the RSS vector measured by TP, the estimated location of TP for KWNN method is as below.

$$\mathbf{D}_j = \| \bar{s}_{l,AP_i} - \bar{s}_{\mathbf{l}_j,AP_i} \| \tag{3}$$

$$\hat{l}_{KWNN} = \frac{\sum_{k=1}^{K} \frac{1}{\mathbf{D}_k} \mathbf{l}_k}{\sum_{k=1}^{K} \frac{1}{\mathbf{D}_k}} \tag{4}$$

$\mathbf{l}_k$ is the known location of *K* nearest RPs and $\mathbf{D}_k$ is the Euclidean distance between the RSS of TP and each of *K* RPs .

2.2.2. A Localization method with considering Altering APs

The existence of altering APs can lead to a low accuracy, since by changing the signals of APs even at the same space, the measured RSS value differs from the previous one in the radio map. AP signals may change with time and make a great influence on localization results. Methods presented in [14,15] are two methods that can improve the results by considering altering APs in indoors.

In these methods, the size of the RSS vector of RPs and TPs must be the same, so when a signal is not sensed from an AP because of variations on signals of APs, a very low value is set for it in the signal vector. The fingerprint vectors of RPs and TPs are divided into several subsets. KWNN algorithm is used to estimate the location of TP in each subset. The k-means algorithm is used to cluster the total estimated locations into the several CHs. Some points of RPs that are closer to each CH are considered. For the test of the similarity of RSS fingerprint vectors in each cluster, cosine similarity is used [16]. The cosine similarity criterion represents the cosine of the angle between the two vectors, which means that when the two angles are very close, their cosine value is higher. The cosine similarity between the fingerprints of the TP and the nearest RPs to each CH in each cluster is performed, and the CH with the highest average of cosine similarity is considered as the estimated location of TP.

$$\cos(\theta) = \frac{\hat{\mathbf{S}}_{\mathbf{1}_j} . \widehat{\mathbf{S}'_l}}{|\hat{\mathbf{S}}_{\mathbf{1}_j}| . |\widehat{\mathbf{S}'_l}|} = \frac{\sum_{i=1}^{AP'} (\bar{\hat{s}}_{\mathbf{1}_j,AP_i} . \bar{\hat{s}}_{l,AP_i})}{\sqrt{\sum_{i=1}^{AP'} \left(\bar{\hat{s}}_{\mathbf{1}_j,AP_i}\right)^2 . \sum_{i=1}^{AP'} \left(\bar{\hat{s}}_{l,AP_i}\right)^2}} \tag{5}$$

where $\hat{\mathbf{S}}_{\mathbf{1}_j}$ is the RSS vector of the j^{th} RP for one of the assumed subsets of APs and $\hat{\boldsymbol{S}}'_l$ is the RSS vector of TP for the same assumed subset of APs. $|\hat{\mathbf{S}}_{\mathbf{1}_j}| . |\hat{\boldsymbol{S}}'_l|$ is the inner product of two RSS vectors of $\hat{\mathbf{S}}_{\mathbf{1}_j}$ and $\hat{\boldsymbol{S}}'_l$. The number of APs in each subset is equal to AP'.

Algorithm in [15] considers the size of the cluster in order to estimate the appropriate cluster of TP, because when the size of the TP cluster becomes large, the estimated location of TP is more accurate. The reason is the issue of increasing the size of the cluster is equal to the presence of none time variant of APs as unchangeable APs. Hence, when there are clusters with lower size, it means that we have outliers that are identified by altering APs. The criteria for each cluster is $\mathbf{I}_C$ as below, that is between 0 and 1.

$$\mathbf{I}_C = \left(\frac{1}{|\mathbf{C}_c|Q}\right)\sum_c \sum_{q=1}^{Q} \cos(\hat{\mathbf{S}}_{q,c}, \hat{\mathbf{S}}_{l_{test,c}}) \tag{6}$$

where Q is the number of the nearest neighbors to each CH that are selected from RPs and $|\mathbf{C}_c|$ is the number of RPs in the c^{th} cluster before adding the Q nearest neighbor RPs to CH of the c^{th} cluster. $\hat{\mathbf{S}}_{q,c}$ is the fingerprint vector of q RPs that are selected as the nearest neighbors to the c^{th} CH. $\hat{\mathbf{S}}_{l_{test},c}$ is the RSS vector of TP that is in cluster c. In this method, when we say that the fingerprint vectors are in the c^{th} cluster, we assume the same APs for it that has been considered in RSS subset samples.

If $|C_{\min}|$ is the smallest cluster size, we can calculate the size of each cluster by $\mathbf{v}_C$,

$$\mathbf{v}_C = e^{-\left(\frac{(|\mathbf{C}_C|-|C_{\min}|)^2}{2b^2}\right)} \tag{7}$$

$\mathbf{v}_C$ is a number between 0 and 1 and b is a parameter for controlling the adjustment of the kernel sensitivity. Next time the score of each cluster is calculated as $\boldsymbol{\xi}_C$.

$$\boldsymbol{\xi}_C = \mathbf{I}_C - \mathbf{v}_C \tag{8}$$

The estimated location of TP, is the CH with the highest value of $\boldsymbol{\xi}_C$. Fig 1 shows an example of methods in [14,15]. The location of a TP is shown by a grey circle. We consider one AP with altering value. The prime estimated locations are clustered to various number of clusters, several RPs adjacent to each of the CHs in terms of the physical Euclidean distance are considered, then we compute the cosine similarity between the RSS of these members of each cluster and the RSS of TP. The CH with the highest average cosine similarity and the biggest size of cluster is the estimated location of TP that is shown in Fig 1 as "True Location" that is marked as a cross mark. The clusters are shown in black circle, black triangle and black pentagram. As shown in Fig 1, the cluster shown by a triangle has unchangeable APs, so the estimated location may be with higher probability at the CH of this cluster.

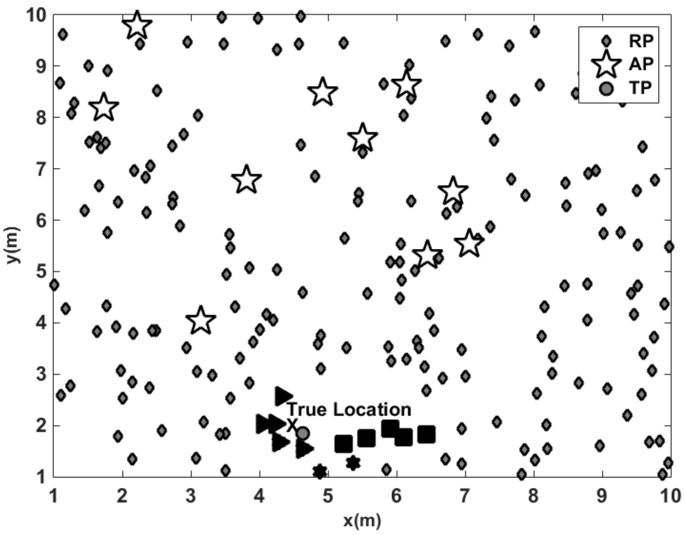

Fig 2 An example of the represented method in section (2.2.2)

2.2.3. Gaussian Kernel Algorithm

It is one of the probabilistic algorithm as a pattern recognition technique that can be used in the location estimation phase. This method uses the Gaussian function $K_{Gauss}(.\,;s_{\mathbf{l}_j,AP_i}(t))$ to calculate the level of similarity between the RSS samples in the radio map and the RSS value of TP as (9). The distribution of $s_{\mathbf{l}_j,AP_i}(t)$ at location $\mathbf{l}_j$ for T seconds can be expressed by a Gaussian kernel model as (10):

$$K_{Gauss}\left(s_{l,AP_i}(t)\ ;s_{\mathbf{l}_j,AP_i}(t)\right) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{\left(s_{l,AP_i}(t)-s_{\mathbf{l}_j,AP_i}(t)\right)^2}{2\delta^2}} \quad (9)$$

$$P(\mathbf{S}_l'|\mathbf{l}_j) = \frac{1}{T}\sum_{t=1}^{T} K_{Gauss}(s_{l,AP_i}(t)\ ;s_{\mathbf{l}_j,AP_i}(t)) \quad (10)$$

δ is the kernel width, that needs to be regulated based on the experiment [28].

2.3. Path-loss Model

We use path-loss model as (11) for simulated radio map [12]. The parameters of the simulated radio map are estimated by measured RSS values. The path-loss formula in our simulations is as below.

$$\mathbf{P}\left(d_{\mathbf{l}_j,AP_i}\right)[dBm] = P\left(d_{0,AP_i}\right)[dBm] - 10n_i \log_{10}\left(\frac{d_{\mathbf{l}_j,AP_i}}{d_{0,AP_i}}\right) + X_{\sigma_i}[dB] - \sum_{j=1}^{w} WAF_j - \sum_{j=1}^{f} FAF_j$$

(11)

The j^{th} RP in distance $d_{\mathbf{l}_j,AP_i}$ from the i^{th} AP can measure $\mathbf{P}\left(d_{\mathbf{l}_j,AP_i}\right)$ as the received power. The received power at 1-meter from the i^{th} AP is $P\left(d_{0,AP_i}\right)$. X_{σ_i} is a zero mean normal distribution with standard deviation σ_i in dB for each AP for shadowing phenomenon. There are two constant values in (11) as wall attenuation factors (WAFs) and floor attenuation factor (FAF) in dB that depend on the material and the number of walls (*w*) and floors (*f*) between an AP and the point. We used the Minimum Mean Square Error (MMSE) fit to the RSS measured data for calculating the path-loss exponent value. Then, σ_i as sigma deviation for each AP is calculated based on the measured signals $\bar{s}_{\mathbf{l}_j,AP_i}$ and the simulated ones as below

$$\sigma_i = \sqrt{\frac{1}{L}\sum_{j=1}^{L}\left(\bar{s}_{\mathbf{l}_j,AP_i} - \mathbf{P}(d_{\mathbf{l}_j,AP_i})\right)} \quad (12)$$

3. Proposed Method

In our proposed method, affinity propagation clustering as a clustering algorithm on signal space in offline phase is used and in location estimation phase, we use a metric that represents the relationship of variance and weight between RSS fingerprint vector of RPs and TP in the cluster where TP belongs to. By these methods we improved the method in section (2.2.2). Also we use a radio map updating algorithm for improving the results of localization by considering altering APs. Our proposed method is shown in Algorithm1.

Algorithm 1
Inputs: RSS vector of RPs, λ, RSS vector of TP, α, z, t_u **Output:** updated radio map, estimated location of TP **Offline phase:** 1) Affinity Propagation Clustering Algorithm on RSS of all RPs (section 3.1.1)

Location estimation phase: 1) Cluster positioning of TP by its RSS vector (section 3.2.1) 2) Analysis on variances of RSSs at each cluster (section 3.2.2) 3) Assigning weight to each point at each cluster (section 3.2.3) 4) Do (section 2.2.2) 5) Estimate the location of TP (section 3.2.4)

Updating the radio map:
Estimate d (section 3.3)
If ($d \leq t_u$)
Update the radio map (section 3.3)
else
Do not update the radio map

3.1. Offline Phase

In this phase we cluster the RSS of RPs by the affinity propagation clustering method.

3.1.1. Affinity Propagation Clustering

Different from k-means clustering algorithm, the idea of affinity propagation algorithm is using the RSS of all RPs and some of them are chosen as CHs [18]. This method chooses CHs better than k-means algorithm. In this method, the time averaged RSS samples of RPs are used. The pairwise similarity between RSS vector of each two RPs is $\mathbf{s}(i,j)$. $\mathbf{s}(i,j)$ is defined as the squared Euclidean distance as shown in (13):

$$\mathbf{s}(i,j) = -\| \mathbf{S}_{\mathbf{l}_j} - \mathbf{S}_{\mathbf{l}_i} \| \qquad i,j = 1, \dots, L \tag{13}$$

There are two types of message transmitted among RPs: 1) responsibility message $\mathbf{r}(i,j)$ transmitted from RPs that are not CH to CH RPs. This message has information about the CHs and 2) availability message $\mathbf{a}(i,j)$ that transmitted from CH RPs to RPs that are not CH. This message is about the attachment relations between RPs and clusters. The i^{th} RP sends the responsibility message to each candidate CH (the j^{th} RP). By this message we can specify the fitness of the j^{th} RP as CH for the i^{th} RP as (14):

$$\mathbf{r}(i,j) \leftarrow \mathbf{s}(i,j) - \max_{j' \neq j}\{\mathbf{a}(i,j') + \mathbf{s}(i,j')\} \tag{14}$$

In the first iteration, $\mathbf{a}(i,j')$ is set to zero. The responsibilities are computed by (14), therefore $\mathbf{r}(i,j)$ is set to the similarity between RPs with index i and j minus the largest similarities between the i^{th} RP and RPs with index j' that are candidate CHs.

The availability message $\mathbf{a}(i,j)$ is sent from each candidate CH to each RP to select RPs in each cluster. This message describes the fitness that the i^{th} RP can be in cluster that the j^{th} RP is its CH:

$$\mathbf{a}(i,j) \leftarrow \min\{0, \mathbf{r}(j,j) + \sum_{i' \neq i,j} \max\{0, \mathbf{r}(i',j)\}\} \tag{15}$$

where $\mathbf{r}(i,i)$ is the self-responsibility that known as the median of input similarities.

$$\mathbf{r}(i,i) = \mathrm{median}\{\mathbf{s}(i,j)\} \tag{16}$$

Self-availability $\mathbf{a}(i,i)$ reflects that the i^{th} RP is a CH based on the positive responsibilities sent from other RPs to the i^{th} RP as a candidate CH:

$$\mathbf{a}(i,i) \leftarrow \sum_{j' \neq i} \max\{0, \mathbf{r}(j',i)\} \tag{17}$$

These two messages are between pairs of RPs with known similarity values. At any RP during affinity propagation method, availabilities and responsibilities are combined to identify the CHs. For the i^{th} RP the value of the k^{th} RP that maximizes $\mathbf{a}(i,k) + \mathbf{r}(i,k)$ defined that the k^{th} RP is CH for the i^{th} RP. The messages are transmitted for a number of iterations.

When the messages are updated for each RP, we must damp them to avoid numerical oscillations that arise in some trends in neighbor RPs. Each message is set to λ times of its value from the previous iteration plus 1-λ times of its updated values. The damping factor λ is between 0 and 1.

3.2. Location Estimation Phase

In this phase, the cluster and the location of TP are estimated.

3.2.1. Cluster Positioning of TP

The signal distances between RSS of CHs and RSS of TP are calculated by NN method [26]. In this deterministic method, by using Euclidean distance between RSS of CHs and TP, we can estimate the cluster of TP:

$$\mathbf{d}_c = \| \bar{s}_{l,AP_i} - \mathbf{C}_{c,AP_i} \| \tag{18}$$

In (18) *i* is the index of APs. $\bar{s}_{l,AP_i}$ is the average RSS measured by TP from the i^{th} AP. $\mathbf{C}_{c,AP_i}$ is the RSS vector of the c^{th} CH from the i^{th} AP. $\mathbf{d}_c$ is the Euclidean distance that has been calculated between signals $\bar{s}_{l,AP_i}$ and $\mathbf{C}_{c,AP_i}$. If we assume that we have *C* number of CHs, the cluster of TP belongs to, is calculated by (19):

$$index_{estimated} = \arg \min_{\substack{c \\ 1 \le c \le C}} \mathbf{d}_c \tag{19}$$

where $index_{estimated}$ is the estimated index of CH where TP belongs to.

3.2.2. Analysis of Variances of RSS at Each Cluster

RSS measurements from different smartphones have different values. However, for better evaluating the location of TP by these alterations in signals, we can detect the trend of RSS of TP. RSS differences between different smart-phones among APs in the same location are more stable than other locations. Therefore, TP can be positioned in a cluster with least variance of RSS differences among APs [17].

TP belongs to the k^{th} cluster. All *N* RPs that measure signals from the i^{th} AP in the k^{th} cluster are $(\mathbf{R}_{k1,AP_i}, \dots, \mathbf{R}_{kN,AP_i})$. *N* is the number of RPs that are in the k^{th} cluster $(n = 1, 2, \dots, N)$. *N* value in each cluster is different from other clusters.

$$\hat{s} = \operatorname*{argmin}_{\substack{n \\ 1 \le n \le N}} \left(\frac{1}{AP}\right) \sum_{i=1}^{AP} \left(\left(\bar{s}_{l,AP_i} - \mathbf{R}_{kn,AP_i} \right) - \bar{\mathbf{r}}_k \right)^2 \tag{20}$$

$$\bar{\mathbf{r}}_k = \frac{\sum_{i=1}^{AP} \left(\bar{s}_{l,AP_i} - \mathbf{R}_{kn,AP_i} \right)}{AP} \tag{21}$$

where AP is the total number of APs. $\bar{\mathbf{r}}_k$ is the mean of differences of RSS values of RPs and RSS of TP over the number of APs in the k^{th} cluster. $\mathbf{R}_{kn,AP_i}$ is the time averaged RSS measured by the n^{th} RP in the k^{th} cluster from the i^{th} AP. The estimated location of TP is $\hat{s}^{th}$ location that is in the k^{th} cluster. Also, this model can improve the results by considering altering APs and localize TP by considering the least variance of RSSs among APs.

3.2.3. Assigning Weight to Points at Each Cluster

We also assign a weight to the points in each cluster which indicates the physical distance is not similar to the signal distance especially in indoor areas, because there are more obstacles than outdoors [19]. Each point in a cluster has similarity with other points in that cluster, therefore for more details, we can assign a weight to each point in the cluster where TP belongs to. This weight represents how much a point in that cluster is similar to TP [19]. The cluster where TP belongs to, is computed based on section (3.2.1). As (11) has shown, the distance between a transmitter and receiver has an exponential relationship with RSS value measured at that point from that transmitter. This criteria can amplify the effect of level of closeness of points to TP by a weighting factor.

$$\mathbf{d}_{wkn} = \sqrt{\sum_{i=1}^{AP} \left(\frac{\mathbf{w}_{ikn}}{W_{kn}}\right) \left(\bar{s}_{l,AP_i} - \mathbf{R}_{kn,AP_i}\right)^2} \tag{22}$$

$$\mathbf{w}_{ikn} = e^{-\left(\frac{\left|\bar{s}_{l,AP_i} - \mathbf{R}_{kn,AP_i}\right|}{z}\right)} \tag{23}$$

where $\mathbf{d}_{wkn}$ is the calculated weighted distance between RSS of TP and RSS of each point in each cluster. w_{ikn} is the weight value of each AP at the n^{th} point that is in the k^{th} cluster. $W_{kn} = \sum_{i=1}^{AP} \mathbf{w}_{ikn}$ is a normalization factor. Therefore $\frac{\mathbf{w}_{ikn}}{W_{kn}}$ is the weight value that assigned to the n^{th} point in the k^{th} cluster for the i^{th} AP. If a point is closer to TP, we can have higher weight for that point since it is more similar to TP than other points. z is a tuning factor. When z tends to infinity, the distance $\mathbf{d}_{wkn}$ goes to Euclidean distance. Therefore, z should be between 0 and 1. A point with the smallest $\mathbf{d}_{wkn}$ is selected as the estimated location of TP. This method can decrease the effects of altering APs on localization results in indoor areas by tuning the z value and get the least weight to these APs, because in the power of the exponential value in (23), the numerator has a high value for farther APs and altering APs and denominator is a parameter between 0 and 1 that can increase the effects of numerator value. Therefore, the weight in (23) becomes a low value and has lower effect on distance calculated in (22) compared with other APs.

3.2.4. Criteria for Location Estimation Phase in Proposed Method

In order to improve the localization results, we estimate the location of the user by the methods represented in section (2.2.2) and in sections (3.2.1-3.2.3). We estimate the location of the user as the mean value of these calculated values.

$$l_{est} = \frac{\mathbf{l}_1 + \frac{\mathbf{l}_2 + \mathbf{l}_3}{2}}{2} \tag{24}$$

where $\mathbf{l}_1$ is the estimated location of user by method in section (2.2.2) and $\mathbf{l}_2$ is the estimated location of user by method in sections (3.2.1- 3.2.2) and $\mathbf{l}_3$ is the location of user in section (3.2.1, 3.2.3). l_{est} is the estimated location of TP. The benefit of this criteria is that, we can get a higher localization accuracy when Wi-Fi signal levels change in indoor areas.

3.3. Updating the Radio Map

The radio map needs to be updated repeatedly for improving the accuracy in localization and this procedure takes a lot of time if the area is surveyed again. Authors in [14] presented a method in which the radio map can be updated by the RSS measured by TP, after the location of TP is estimated. RSS of a TP captured the changes in the real environment that can be valuable. The measured RSS along with the TP location can update the radio map.

First, the distance between the estimated location of TP and the location of the nearest RP is calculated as d value. We set a threshold t_u (a few meter) for RSS values to not change too much. By considering $d \leq t_u$, we can expect the RSS value of the nearest RP changes due to the effects of environmental dynamicity on signals. For updating the radio map, a parameter α is used which is between 0 and 1.

$$\widehat{\mathbf{V}} \leftarrow (1-\alpha)\widehat{\mathbf{V}} + \alpha\widetilde{\mathbf{V}} \tag{25}$$

where $\widehat{\mathbf{V}}$ is the RSS value of RP that the estimated location of TP is closer to the location of this RP. $\widetilde{\mathbf{V}}$ is the RSS value measured by TP. Therefore, the RSS value measured by RP that is the nearest point to the estimated location of TP is updated to a new value $\widehat{\mathbf{V}}$ and if $d > t_u$ the radio map is not updated.

4. Experimental Setup and Simulation Parameters

The Cyberspace Research Institute at Shahid Beheshti University is our layout for simulation where we sensed 4 APs in this area. Fig 2 shows this layout. Our measuring device in offline and online phases is Samsung Galaxy Grand Prime Smartphone. The size of the considered area is 36.88 m in x-axis and 9.28 m in y-axis. We measured RSS vector of 164 RPs and 20 TPs in this area. The number of measured RSS time samples in each location is 80. Among sensed APs in our layout, AP2 is positioned in downstairs. We measured -112dBm in some points especially in rooms when we cannot measure any signal from some APs. Measured RSS data for sensed APs in our layout are used in our simulations also for higher number of APs, the simulated radio map is used.

4.1. Building the Simulated Radio Map

In order to calculate the simulation data, we first consider the whole area in a gridding board and the wall equations and equations of each two points are calculated by the crossings of these two equations with each other. By this method, the attenuations of walls by considering the thicknesses of each wall are considered. We estimate the sigma deviation for path-loss formula for our sensed APs using (12). The average value of correlation of mean RSS values of points from measurements and simulated data in all points from all 4 sensed APs is 0.89. This correlation value shows that our method for estimating the parameters of path-loss formula and WAFs and FAF are almost accurate. Therefore, we use these parameters for simulated radio maps.

Path-loss exponent values and sigma deviations for each AP, in radio maps that have beyond 4 APs are computed by (26) and (27). We use 4, 10 and 18 APs in the whole area for our simulations. We fixed 4 APs that are in our layout in all radio maps. We set the power of transmitters equal to 27dBm. We considered -30dB as floor attenuation factor for AP2. The path-loss exponent value for AP1, AP2, AP3 and AP4 that are shown in Fig 2 are respectively equal to 5.59, 6.92, 5.53 and 6.09. The sigma deviation for these APs are computed respectively equal to 11.74, 14.79, 11.84 and 10.30dB. The path-loss exponent value for all APs in all radio maps are calculated by path-loss exponent values of these four APs. Also we use this criteria for calculating the sigma deviation for all APs in all radio maps:

$$n_i = (1 - \beta)n_{AP(WAF)} + (\beta)n_{AP(loc)} \tag{26}$$

$$\sigma_i = (1 - \beta)\sigma_{AP(WAF)} + (\beta)\sigma_{AP(loc)} \tag{27}$$

where β is a parameter between 0 ,1. $n_{AP(loc)}$ is the path-loss exponent value of AP that is closer to AP_i in the case of location. $n_{AP(WAF)}$ is the path-loss exponent value of AP that has closer WAF value to AP_i. $\sigma_{AP(loc)}$ is the sigma value of AP that is closer to AP_i in the case of location. $\sigma_{AP(WAF)}$ is the sigma value of AP that has closer WAF value to AP_i.

We use β=0.5 for our simulations. We use this method for simulation parameters of APs in the radio maps that have beyond our 4 sensed APs. As mentioned, the simulation parameters for simulated radio map depend on RSS measured. Therefore, when we do not have any RSS measured values from some APs for simulation, it is better to consider their simulation parameters by the help of the parameters of the closest AP to them. In our consideration, we assumed that when two APs have closer WAF value are closer to each other in the case of space, they have nearly equal path-loss exponent and sigma deviation values too.

We use the results of six nearest neighbours for KNN and δ=0.25dB for Gaussian kernel methods. We use six weighted nearest neighbours for KWNN and b equal to 0.25 for method in section (2.2.2). In the new proposed method, we set the number of CHs in offline phase for 4, 10 and 18 APs equal to 14 that are calculated by affinity propagation. Also we set λ=0.5 in simulations as damping factor for affinity propagation clustering algorithm.

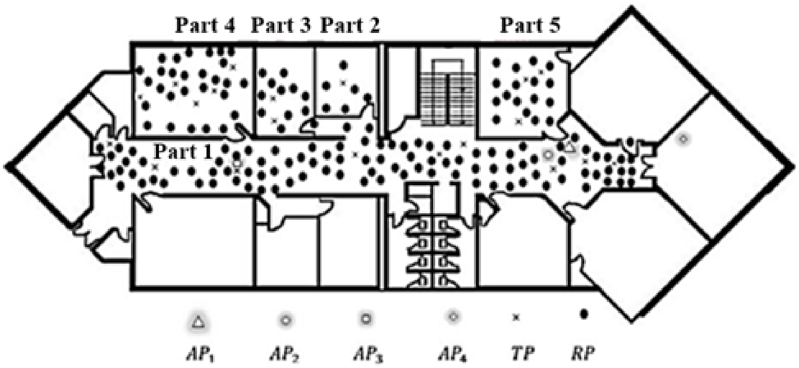

Fig 3 Layout of Cyberspace Research Institute at Shahid Beheshti University

Altering APs decreases the accuracy of localization in reality, however, we want to simulate our proposed algorithm in simulation. Therefore, we should consider Signal Change value as a parameter that represents how much the power of altering APs differs when considering their variations on signals in location estimation phase. We consider 0 to 20dBm as Signal Change values. Fig 3 shows the mean distance error versus Signal Change values for only one altering AP when we have 4 APs.

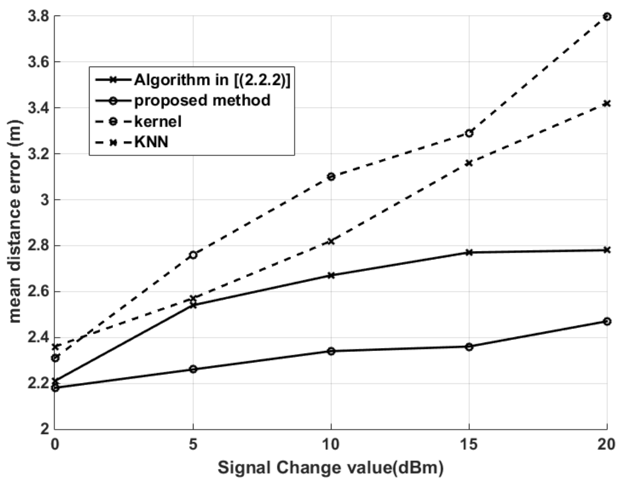

Fig 4 Mean distance error versus Signal Change values

As shown in Fig 3, mean distance error for the proposed method is lower than other methods, because our proposed method uses RSS subset sampling for localization as mentioned in section (2.2.2) and also the variability of variance and weights of points that are in cluster where TP belongs to, are analysed. The trend of increasing the mean distance error by increasing the Signal Change value for methods are nearly the same. Mean distance error for kernel and KNN methods are higher than method represented in section (2.2.2) because these methods can not consider altering APs. When we have Signal Change value equal to the range of 0 to 20dBm, the signal coverage area for altering APs is increased therefore the mean distance error in these ranges is increased for all methods.

We consider Signal Change value of 10dBm for altering APs for the simulations throughout this paper. In section (3.2.3) we have z parameter that can be regulated between 0 and 1. The mean distance error for various values of z is computed. Fig 4 shows these results when there are 4 APs in radio map by considering one altering AP. As shown in this figure, mean distance error in our proposed method has nearly increasing trend by increasing the z value.

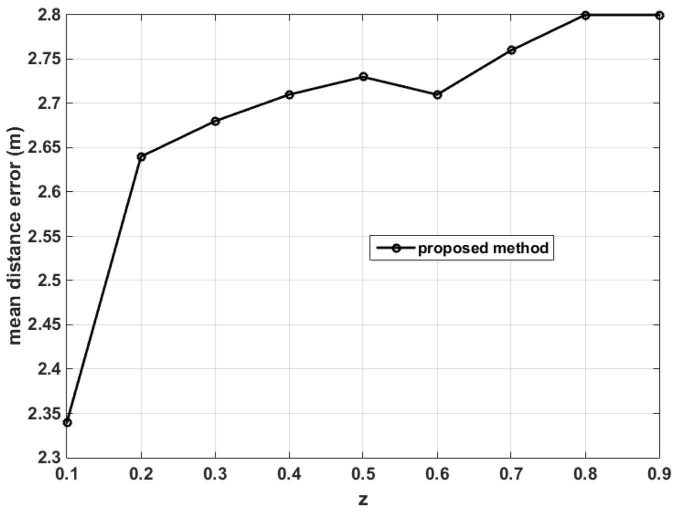

Fig 5 Mean distance error for various *z* values

We use *z*= 0.1 in our simulations. The *z* value can better relate the relationship between signal distance and physical distance, because the criteria of comparison for signal distance is not equal to the Euclidean distance. Therefore, by increasing the *z* value, the distance between signals tend to Euclidean distance.

5. Simulation Results

For analysis of simulation of our proposed method, we compare it with KNN and kernel methods and the method represented in section (2.2.2).

We analysis the results of measured and simulated radio map with 4 sensed APs in our layout when we have one altering AP in simulated and measured radio map then we compare it when we have these APs without any altering AP in measured radio map. The results of these three scenarios are shown in Fig 5.

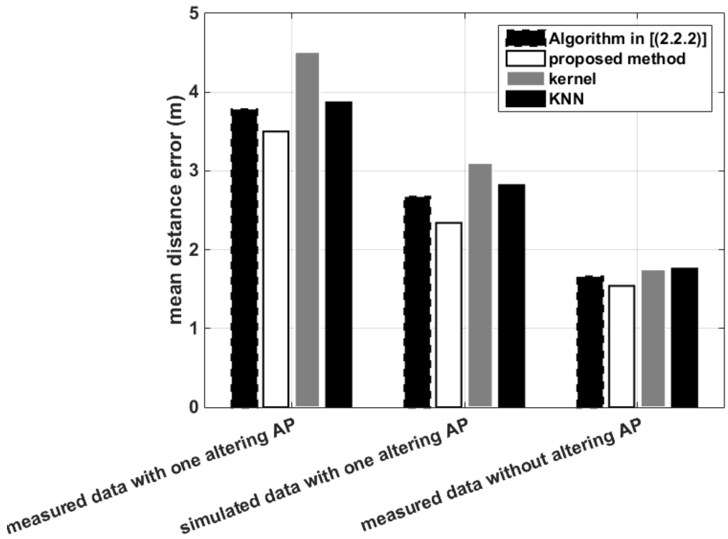

Fig 6 Mean distance error for three scenarios by 4 APs

As shown in Fig 5, our proposed method has the best accuracy in all three scenarios and when we have one altering AP, the results of localization have lower accuracy compared with no altering AP. We cannot measure any signal in some where especially in rooms from some APs and the results of measured data have lower accuracy than simulated data by considering altering APs. Therefore, when we use measured data with no altering AP, errors of algorithm represented in section (2.2.2), our proposed method, KNN and kernel methods are equal to 1.66, 1.54, 1.76 and 1.75m, respectively, that are nearly close to each other, because in reality we do not consider any RSS data in rooms from some Aps. Therefore, we should expect to have close results in all three methods in reality, because we do not have much more RSS variances.

Table 1 shows the mean distance error for various number of altering APs when we have radio map with 18 dimensions and we use one, five and ten altering APs. As shown in this Table, mean distance error in all methods is increased by increasing the number of altering APs. However, the proposed method has lower error than other methods because it considers both clustering the RSS vectors and analysis of variance of RSS value and signal distance more deeply compared with others and clustering the locations by RSS subset sampling that is better in learning in dynamic effects in indoor environments.

Table 1 mean distance error (m)

Method / Number of altering APs	1	5	10
Algorithm in (2.2.2)	1.86	2.70	4.02
proposed method	**1.47**	**2.59**	**3.7**
kernel	1.98	2.81	4.36
KNN	1.92	2.78	4.28

According to Table 1, the mean distance error increases when the number of altering APs increases. For simplicity, we evaluate the proposed method when there is only one altering AP for the rest of the paper. The localization results of all 164 RPs are shown in Fig 6 by considering one altering AP. As shown in Fig 6, by increasing the number of APs, the mean distance error decreases and the proposed method has shown more improvement than other methods. Because the RSS subset sampling (section 2.2.2) and the RSS variances among APs in each cluster in location estimation phase are computed and also we assigned a weight to each point in each cluster, that can better represent the relationship between signal distance and physical distance. The reason is that in indoor areas, there are obstacles that may change the AP signals.

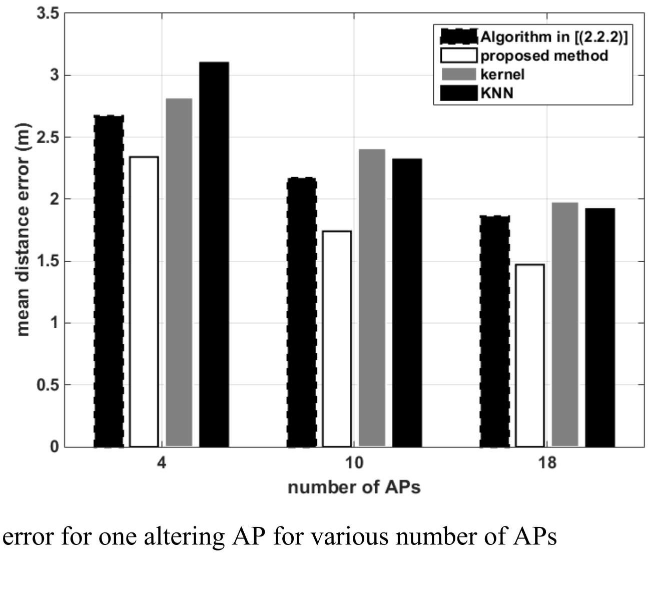

Fig 7 Mean distance error for one altering AP for various number of APs

By these simulation results, we find out that for 4, 10 and 18 APs with only one altering AP and 164 RPs, the results of method in section (2.2.2) are improved by 12%, 20% and 21% that in average is 17.67%. As simulation results have shown, when we have lower numbers of unchangeable APs, the errors in localization results become higher.

We can update the radio map as we discussed in section (3.3). We set t_u=1.3 m as the threshold distance value and we simulate all methods for various α values as shown in Fig 7 for 4 APs with one altering AP.

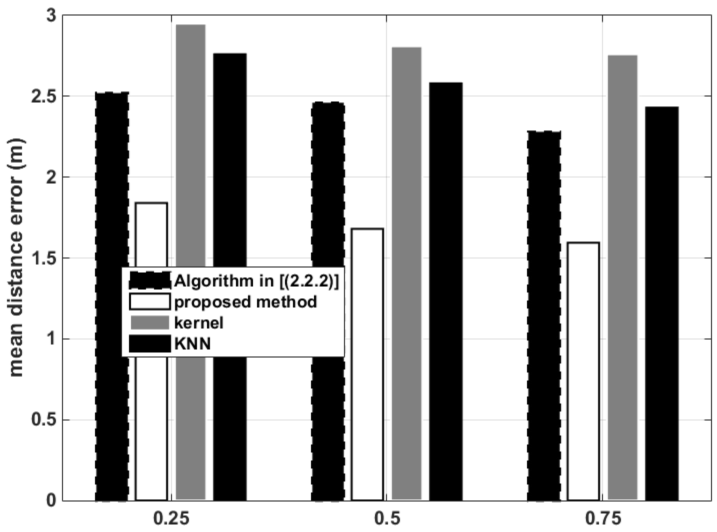

Fig 8 Mean distance error by various α values for 4 APs with one altering AP

As shown in Fig 7, when α=0.75 there are more improvements in localization results in all methods. The reason is that in this case, the RSS value of TP is used more than RSS value of RP for updating the radio map. Therefore, it is normally possible that the results of localization be much more improved. However, for α=0.5 the RSS of TP and RP have the same share. This situation is more common in reality because the RSS value has non-stationary values with time, therefore it should be better that we consider this case for our simulations. Thus, we select α=0.5 for updating the radio maps with one altering AP. The results of this analysis are shown in Fig 8.

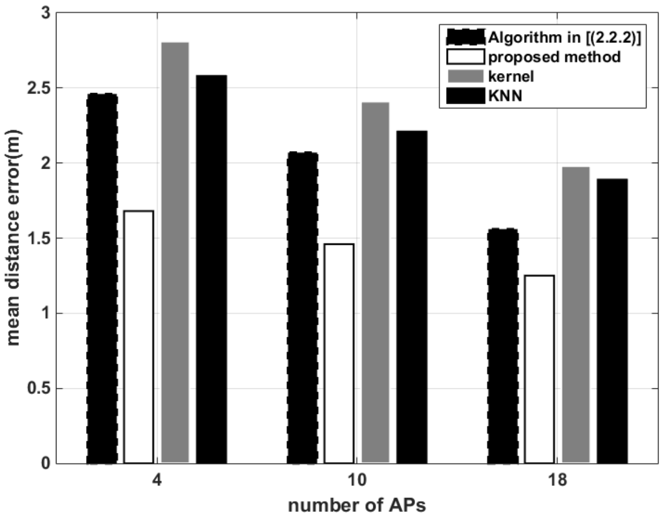

Fig 9 Mean distance error by α=0.5 for all radio maps with one altering AP

As shown in Fig 8, mean distance error decreases by increasing the number of unchangeable Aps. Also, in average we improve the results of method in section (2.2.2) as 22.5% when we have one altering AP by this radio map updating algorithm. The results have shown that our proposed method can filter altering APs and decreasing their effects on localization results, without deleting them from our simulations. The improvement on KNN and kernel methods by this radio map updating algorithm in average is equal to 8% and 16%, respectively, and the proposed method has more improvements.

6. Conclusion

The proposed method in this paper considers the dynamicity on Wi-Fi signals such as altering APs and improves the results of localization without deleting altering APs. Our proposed method uses affinity propagation clustering algorithm in signal space and it has considered the least variance of RSS measurements from APs in each cluster in order to get higher accuracy for location estimation of TP. Also we assign lower weight to altering APs in each cluster that can better represent the exponential relationship between physical distance and signal distance in indoors. Also, we use an algorithm for updating the radio map. Then we get in average 22.5% improvement in compare of when we simulate RSS subset sampling for localization because of altering APs.

Acknowledgements Vahideh Moghtadaiee gratefully acknowledges the support provided by the Iran National Science Foundation (INSF) for this work.

REFERENCES

1. He, S., & Chan, S.-H. G. (2016). Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials, 18*(1), 466-490.
2. Munoz, D., Lara, F. B., Vargas, C., & Enriquez-Caldera, R. (2009). *Position location techniques and applications*: Academic Press.
3. Harroud, H., Ahmed, M., & Karmouch, A. (2003). Policy-driven personalized multimedia services for mobile users. *IEEE Transactions on Mobile computing, 2*(1), 16-24.
4. Molina, B., Olivares, E., Palau, C. E., & Esteve, M. (2018). A Multimodal Fingerprint-Based Indoor Positioning System for Airports. *IEEE Access, 6*, 10092-10106.
5. Rodriguez, M. D., Favela, J., Martínez, E. A., & Muñoz, M. A. (2004). Location-aware access to hospital information and services. *IEEE Transactions on information technology in biomedicine, 8*(4), 448-455.
6. Misra, P., & Enge, P. (1999). Special issue on global positioning system. *Proceedings of the IEEE, 87*(1), 3-15.
7. Raquet, J., & Martin, R. K. (2008). Non-GNSS radio frequency navigation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5308-5311).
8. Vo, Q. D., & De, P. (2016). A survey of fingerprint-based outdoor localization. *IEEE Communications Surveys & Tutorials, 18*(1), 491-506.
9. Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An in-building RF-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings.* (pp. 775-784).
10. Moghtadaiee, V., & Dempster, A. G. (2014). Indoor location fingerprinting using FM radio signals. *IEEE Transactions on Broadcasting, 60*(2), 336-346.
11. Liu, Y., Dashti, M., Rahman, M. A. A., & Zhang, J. (2014). Indoor localization using smartphone inertial sensors. In *Positioning, Navigation and Communication (WPNC)*. (pp. 1-6).
12. Alikhani, N., Amirinanloo, S., Moghtadaiee, V., & Ghorashi, S. A. (2017). Fast fingerprinting based indoor localization by Wi-Fi signals. In *International Conference on Computer and Knowledge Engineering (ICCKE)*. (pp. 241-246).
13. Khalajmehrabadi, A., Gatsis, N., & Akopian, D. (2017). Modern WLAN fingerprinting indoor positioning methods and deployment challenges. *IEEE Communications Surveys & Tutorials, 19*(3), 1974-2002.
14. He, S., Ji, B., & Chan, S.-H. G. (2016). Chameleon: Survey-free updating of a fingerprint database for indoor localization. *IEEE Pervasive Computing, 15*(4), 66-75.
15. He, S., Lin, W., & Chan, S.-H. G. (2017). Indoor localization and automatic fingerprint update with altered AP signals. *IEEE Transactions on Mobile Computing, 16*(7), 1897-1910.
16. Han, S., Zhao, C., Meng, W., & Li, C. (2015). Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity. In *International Conference on*

Communications (ICC). (pp. 2710-2714).
17. Wang, B., Chen, Q., Yang, L. T., & Chao, H.-C. (2016). Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches. *IEEE Wireless Communications, 23*(3), 82-89.
18. Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *science, 315*(5814), 972-976.
19. Fu, Y., Chen, P., Yang, S., & Tang, J. (2018). An Indoor Localization Algorithm Based on Continuous Feature Scaling and Outlier Deleting. *IEEE Internet of Things Journal, 5*(2), 1108-1115.
20. Huang, J., Millman, D., Quigley, M., Stavens, D., Thrun, S., & Aggarwal, A. (2011). Efficient, generalized indoor wifi graphslam. In *International Conference on Robotics and Automation (ICRA).* (pp. 1038-1043).
21. Yoo, S., Kim, E., & Kim, H. (2014). Exploiting user movement direction and hidden access point for smartphone localization. *Wireless personal communications, 78*(4), 1863-1878.
22. Ferris, B. D., Fox, D., & Lawrence, N. (2007). WiFi-SLAM using Gaussian process latent variable models.
23. Yin, J., Yang, Q., & Ni, L. M. (2008). Learning adaptive temporal radio maps for signal-strength-based location estimation. *IEEE transactions on mobile computing, 7*(7), 869-883.
24. Atia, M. M., Noureldin, A., & Korenberg, M. J. (2013). Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Transactions on Mobile Computing, 12*(9), 1774-1787.
25. Fallah, M., & Pourahmadi, V. (2018). Graph-based iterative measurement-denoising and radio-map generation for semi-supervised indoor localisation. *IET Communications, 12*(7), 848-853.
26. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*: MIT press.
27. Roos, T., Myllymäki, P., Tirri, H., Misikangas, P., & Sievänen, J. (2002). A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks, 9*(3), 155-164.
28. Kushki, A., Plataniotis, K. N., & Venetsanopoulos, A. N. (2007). Kernel-based positioning in wireless local area networks. *IEEE transactions on mobile computing,* (6), 689-705.