RESEARCH　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Fingerprint indoor positioning algorithm based on affinity propagation clustering

Zengshan Tian, Xiaomou Tang[*], Mu Zhou and Zuohong Tan

Abstract

Recently, the fingerprint-based wireless local area network (WLAN) positioning has gained significant interest. A probability distribution-aided indoor positioning algorithm based on the affinity propagation clustering is proposed. Different from the conventional fingerprint-based WLAN positioning algorithms, the paper first utilizes the affinity propagation clustering to minimize the searching space of reference points (RPs). Then, we introduce the probability distribution-aided positioning algorithm to obtain the target's refined position. Furthermore, because the affinity clustering can effectively lead to a reduction of the computational cost for the RP searching which is involved in the probability distribution-aided positioning algorithm, the proposed algorithm can lower the difficulty and minimize the power consumption when estimating the user's position. Experimental results conducted in the real environments show that our proposed algorithm will significantly improve the performance of the probability distribution-aided positioning algorithm in both the positioning accuracy and real-time ability.

Keywords: WLAN indoor positioning; Fingerprinting; Affinity propagation clustering; RSS; Probability distribution

1 Introduction

In recent decade, the indoor wireless local area network (WLAN) positioning technology has caught significant attention by a variety of universities and research institutes [1-3]. Among them, the time of arrival (ToA), angle of arrival (AoA), and received signal strength (RSS) are the three most representative measurements for the position estimation. Compared to the ToA and AoA measurements, the RSS can be more easily measured without any additional special hardware devices in current open public WLAN networks. However, the most significant challenge of the RSS readings is about the irregular variations of RSS due to the variable radio channel attenuation, signal shadowing, multipath interference, and even the variations of indoor temperature [4].

In general, two approaches are used by the existing WLAN positioning techniques for position estimation. One effective solution is the k-nearest neighbor (kNN) algorithm to estimate the mobile user's position at the centroid of the K closest neighbors. The closest neighbors are defined as the reference points (RPs) which have the smallest RSS distance to the on-line new collected RSS readings [5]. The kNN algorithm can be easily implemented by the current widely existing WLAN infrastructures, while the accuracy is limited. Another alternative approach is based on the statistical analysis on the probabilities of each candidate RP to calculate the confidence probability of each RP to be selected as the mobile user's estimated position [6,7].

RSS clustering by the measures of similarities between the RSS readings can be suggested as a critical step for the fingerprint-based WLAN positioning. Performing the RSS clustering prior to the positioning process has two main advantages. First, it helps to mitigate the degradation on the positioning accuracy caused by the RSS deviations and potential outlier readings. Second, it is beneficial to be used to reduce the computation cost since only the cluster centers are considered for the positioning [8].

In this paper, we present a new accurate and scalable positioning algorithm to estimate the user's position with low computation cost in a public WLAN environment. Our algorithm consists of two steps: (1) the coarse positioning step is used to obtain the cluster which the

* Correspondence: tangxiaomou@foxmail.com
Chongqing Key Lab of Mobile Communications Technology, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Springer

user belongs to; and (2) the fine positioning step is utilized to calculate the accurate coordinates of the user.

The paper is organized as follows. Section 2 provides some related work on the fingerprint indoor positioning algorithms. Section 3 discusses the overall structure of our algorithm. Sections 4 addresses the detailed steps of the off-line affinity propagation, on-line cluster matching-based coarse positioning, and probability distribution-aided fine positioning, respectively. The performance of our proposed algorithm is verified in Section 5. Finally, Section 6 concludes this paper.

2 Related work

The RSS fingerprints are recognized as the vectors of RSS values recorded from the hearable access points (APs) in target area. The fingerprint indoor positioning by using the set of pre-calibrated RSS fingerprints (or called the radio map) can be normally classified into three categories: the (1) deterministic approach; (2) probabilistic approach; and (3) machine learning approach, as illustrated in Figure 1. The widely known RADAR system [9,10] is addressed as one of most representative deterministic approaches for the WLAN fingerprint indoor positioning by using the K-nearest neighborhood RPs to infer the user's position.

The basic idea of probabilistic approach is to pre-store the RSS distribution with respect to each hearable AP into a radio map and use it to conduct the position estimation. As an example, the Horus system [11] estimates the user's position at the location which has the largest posterior probability by the Bayesian inference [12]. Although the Bayesian inference achieves high positioning accuracy, the recording of RSS distributions at RPs is time consuming.

Due to the significant computation cost involved in the previously mentioned approaches, the researchers begin to pay more attention to the machine learning approach (e.g., the artificial neural network [13,14], support vector machine [15], and fuzzy logic [16,17]) to realize the fingerprint positioning. The most important advantage of the machine learning approach is about the real-time ability of inferring the user's coordinates in the on-line phase. However, most of the current machine learning positioning systems is designed for a small-scale

area (e.g., approximately 600 m^2 [18]), and meanwhile, the positioning accuracy will significantly rely on the training process in the off-line phase.

To improve the positioning accuracy further, we always consider about the fingerprint processing, such as the RSS clustering and dimension reduction. A conventional way to the RSS clustering is to select a set of cluster centers to minimize the sum of the squared distances between the RSS readings and their corresponding centers. As a representative, the widely used k-means clustering begins with an initial set of randomly selected centers and then iteratively refines this set to decrease the sum of the squares distances. However, the k-means clustering is quite sensitive to the initial selection of centers. In this case, it is always performed in several times and with different initial centers to find the best clustering results. Therefore, the k-means clustering is limited in practical use due to the arbitrary selection of the initial cluster centers. To solve this problem, the affinity propagation clustering creates the centers and the corresponding clusters based on the constant exchanging of reading similarities between the RPs [19]. With this idea, the messages will be exchanged between RPs until a high-quality set of centers and corresponding clusters gradually emerge.

3 Overall structure of proposed positioning algorithm

The block diagram of our proposed indoor positioning algorithm is shown in Figure 2. Obviously, this algorithm contains two phases: (1) in the off-line phase, we construct the radio map and conduct the affinity propagation clustering; and (2) in the on-line phase, the cluster matching-based coarse and probability distribution-aided fine positioning will be performed, respectively. The detailed steps about the block diagram in Figure 2 will be analyzed in Section 3.

In the off-line phase, we first hold the WLAN mobile device to collect the RSS readings from the hearable APs to construct the radio map in the area of interest. During the construction of the radio map Ψ (see details below), the corresponding physical coordinates of RPs should also be stored. Then, the affinity propagation
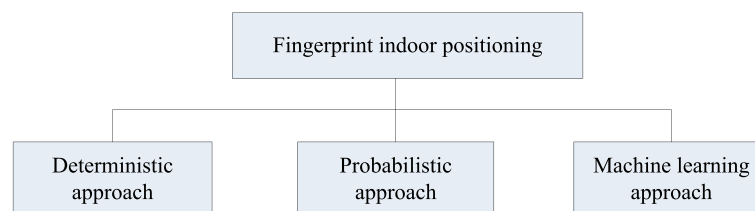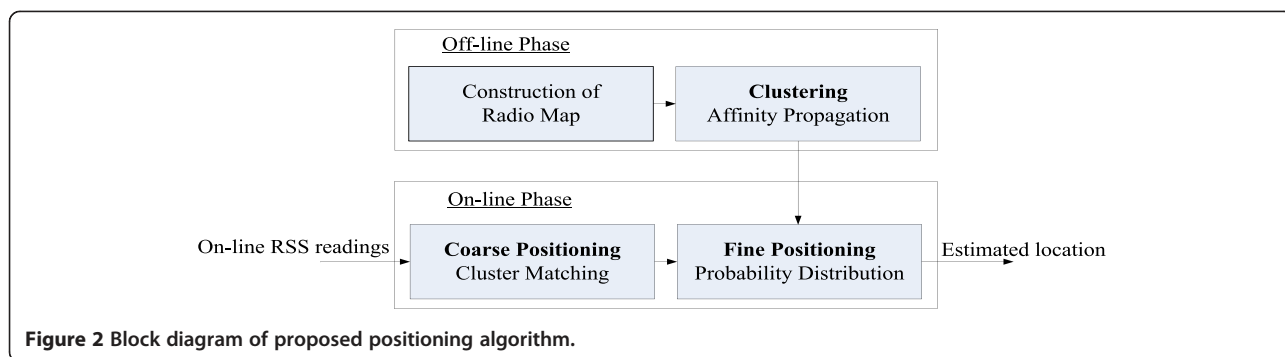

Figure 1 Categories of fingerprint indoor positioning.

Figure 2 Block diagram of proposed positioning algorithm.

clustering is conducted in the radio map to cluster the raw RPs.

In the on-line phase, the first step is to use the mobile device to collect the on-line new RSS readings. Then, the coarse positioning will be used before the fine positioning for the reasons of reducing the on-line computation cost and improving the accuracy performance of the probability distribution-aided positioning algorithm.

Due to the time-variation property of radio propagation in indoor WLAN environment (e.g., multi-path effect, RSS shadowing, and adjacent channel interference), the fingerprint-based positioning has been more preferred in practical use [20]. During the off-line phase, the RSS readings are collected at pre-calibrated positions which are also named as the RPs. We denote the τ – th RSS readings from AP_i at RP_j as $\{\psi_{i,j}(\tau), \tau = 1, ..., q, q > 1\}$ where q is the number of RSS readings. Normally, the average of RSS readings will be computed and stored into a database which is known as the widely recognized radio map Ψ. The radio map can effectively describe the spatial distribution property of the RSS in target area.

$$\Psi = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,N} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{L,1} & \psi_{L,2} & \cdots & \psi_{L,N} \end{pmatrix}, \tag{1}$$

where $\psi_{i,j} = \frac{1}{q}\Sigma_{\tau=1}^{q}\psi_{i,j}(\tau)(i = 1, 2, ..., L; j = 1, 2, ..., N)$ is the average of RSS readings from AP_i at RP_j over the time domain. L and N are the number of access points (APs) and RPs, respectively. Each column of the radio map Ψ (i.e., $\vec{\psi_j} = \left[\psi_{1,j}, \psi_{2,j}, ..., \psi_{L,j}\right]^T, j = 1, 2, ..., N$) represents a sequence of RSS readings at a RP_j. The superscript 'T' denotes the transposition operation.

4 Detailed steps of proposed positioning algorithm

Different from the conventional K-means clustering [21], the basic idea of our proposed affinity propagation clustering algorithm is to use the *preference* (p) to label the RPs and the RPs with larger preference are more likely to be selected as the cluster centers. Our proposed algorithm outperforms the K-means clustering because of the initialization-independent property and better selection of cluster centers [22].

For the affinity propagation clustering, we first use the pairwise similarity $s(i, j)$ to describe the fitness of the RP_j to be selected as the cluster center with respect to the RP_i. Based on Equation (1), we can denote the RSS vector for each RP_j as $\vec{\psi_j} + \vec{\delta_j}$ where $\vec{\delta_j}$ is the measurement noise which obeys the Gaussian distribution. Therefore, the pairwise similarity $s(i, j)$ can be defined as the squared Euclidean distance in Equation (2).
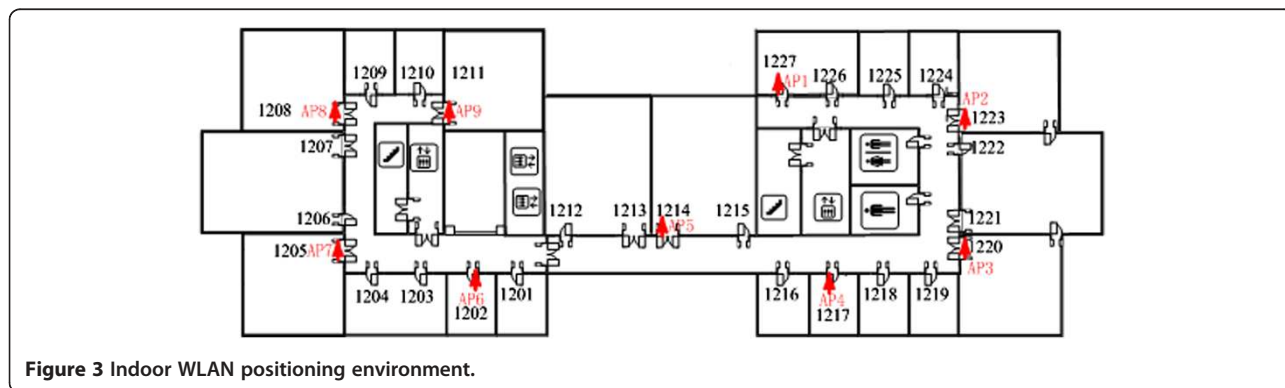

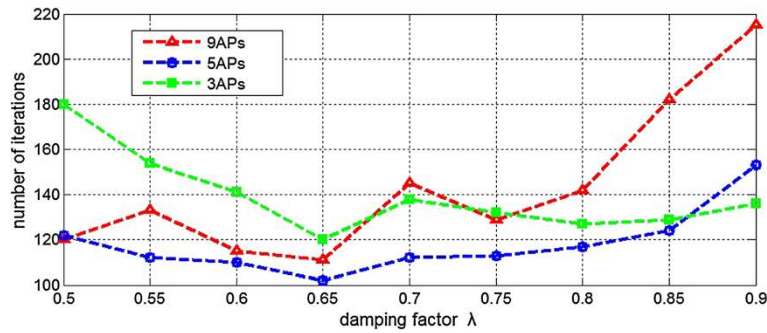**Figure 3 Indoor WLAN positioning environment.**

Figure 4 Variations of the number of iterations with respect to the damping factor.

$$s(i,j) = -\left\| \vec{\psi}_i - \vec{\psi}_j \right\|^2, \forall i,j \in \{1,2,...,N\} \quad (2)$$

Furthermore, there are two types of messages transmitted among the RPs for the affinity propagation clustering: (1) responsible message $r(i,j)$ which transmits the information about the clustering center; and (2) availability message $a(i,j)$ which informs the attachment relations between the RPs and clusters.

The RP_i will send the responsible message to each candidate cluster center RP_j to transmit the accumulated fitness for the RP_j to be selected as the cluster center for RP_i. By taking all the other potential cluster centers j' for RP_i into account, we can obtain

$$r(i,j) = s(i,j) - \max_{j' \neq j}\{a(i,j') + s(i,j')\}, \quad (3)$$

where $a(i,j)$ is the availability message, as defined in Equation (5). Meanwhile, we define the self-responsibility $r(i,i)$, which is known as *preference (p)* as the median of input similarities, resulting in the average number of clusters.

$$p = \text{median}\{s(i,j), \forall i,j \in \{1,2,...,N\}\} \quad (4)$$

The availability message $a(i,j)$ is sent from each candidate cluster center RP_j to RP_i. $a(i,j)$ describes the accumulated fitness for RP_i to select RP_i as its center, such as

$$a(i,j) = \min\left\{0, r(j,j) + \sum_{i' \neq i,j} \max\{0, r(i',j)\}\right\}. \quad (5)$$

Similarly, the self-availability $a(j,j)$ will reflect the accumulated fitness for RP_j to be selected as the center. Because of the requirement of positive responsibilities, we have

$$a(j,j) = \sum_{i' \neq j} \max\{0, r(i',j)\}. \quad (6)$$

The previously mentioned messages are transmitted among the neighboring RPs until the optimal cluster centers are searched out. When updating the messages, it is important that they be damped to avoid numerical oscillations that arise in some circumstances. Each message is set to λ times its value from the previous iteration plus $1 - \lambda$ times its prescribed updated value, where
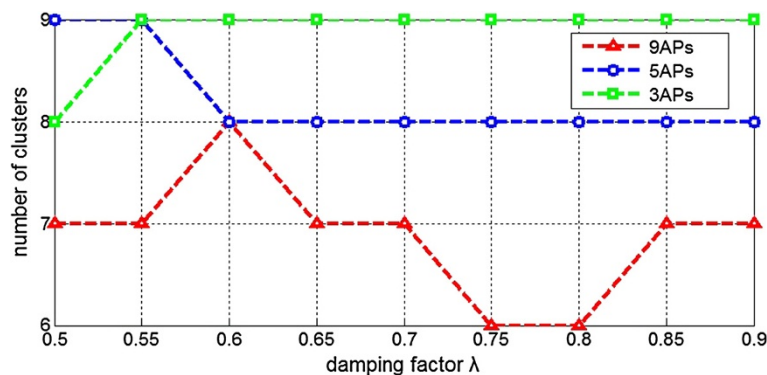

Figure 5 Variations of the number of clusters with respect to the damping factor.
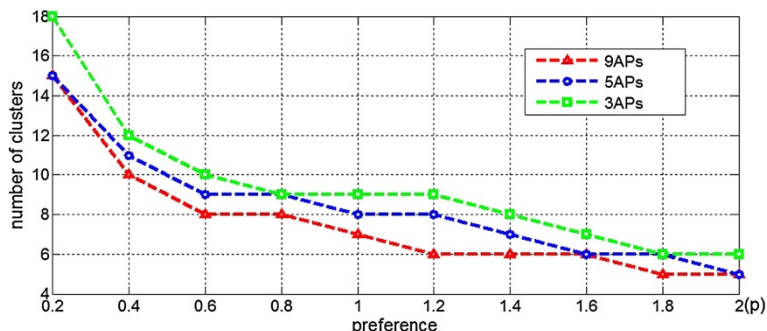
Figure 6 Variations of the number of clusters with respect to the parameter p.

the damping factor λ is between 0 and 1. In each iteration, there are three steps involved [23]: (1) updating the responsibilities by Equation (7); (2) updating the availabilities by Equation (8); and (3) integrating the availabilities and responsibilities to determine the cluster centers by Equation (9). In our experiments, we keep on this process until the cluster centers have not changed in ten iterations or the number of iterations exceeds 300.

$$r_i = (1-\lambda)r_i + \lambda r_{i-1} \qquad (7)$$

$$a_i = (1-\lambda)a_i + \lambda a_{i-1} \qquad (8)$$

$$r(i,i) + a(i,i) > 0 \qquad (9)$$

In the on-line phase, we will collect the new RSS readings at unknown positions $\vec{\psi_r} = \left[\psi_{1,r}, ..., \psi_{L,r}\right]^T$ where $\{\psi_{k,r}, k = 1, ..., L\}$ is the average of new RSS readings from AP_k. We define H and C_j as the set of cluster centers and the set of RPs with the center $RP_j \in H$. After the coarse positioning, the candidate cluster selected for the fine positioning can be obtained by Equation (10).

$$j' = \arg\min_{j \in H} || \vec{\psi_r} - \vec{\psi_j} ||^2, \vec{\psi_r} = \left[\psi_{1,r}, ..., \psi_{L,r}\right]^T \qquad (10)$$

For the fine positioning, we will calculate the matching probability between the on-line new collected RSS readings and the pre-stored fingerprints in radio map. By assuming the Gaussian probability distribution of RSS readings at each RP, the RSS values should obey the normal distribution $N(\mu, \sigma^2)$ [7].

First, the likelihood function is calculated by

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \qquad (11)$$

Second, we can obtain the logarithmic equation in Equation (12).

$$\lg L(\mu, \sigma^2) = -\frac{n}{2}\lg(2n) - \frac{n}{2}\lg(\sigma^2) - \frac{n}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad (12)$$

Third, likelihood equations should be

$$\begin{cases} \dfrac{\partial \lg L(\mu, \sigma^2)}{\partial \mu} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \\[2ex] \dfrac{\partial \lg L(\mu, \sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases}. \qquad (13)$$
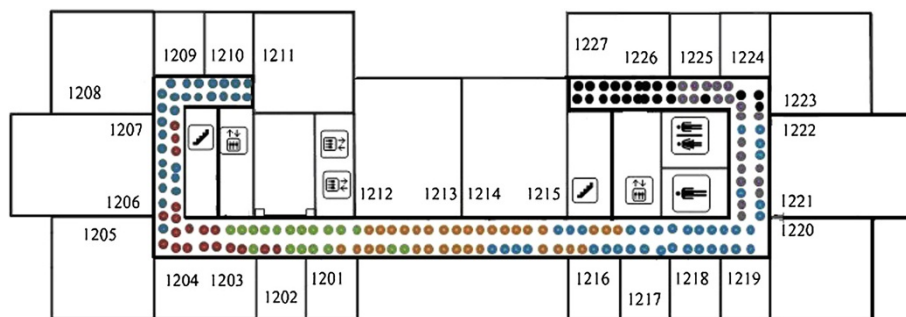

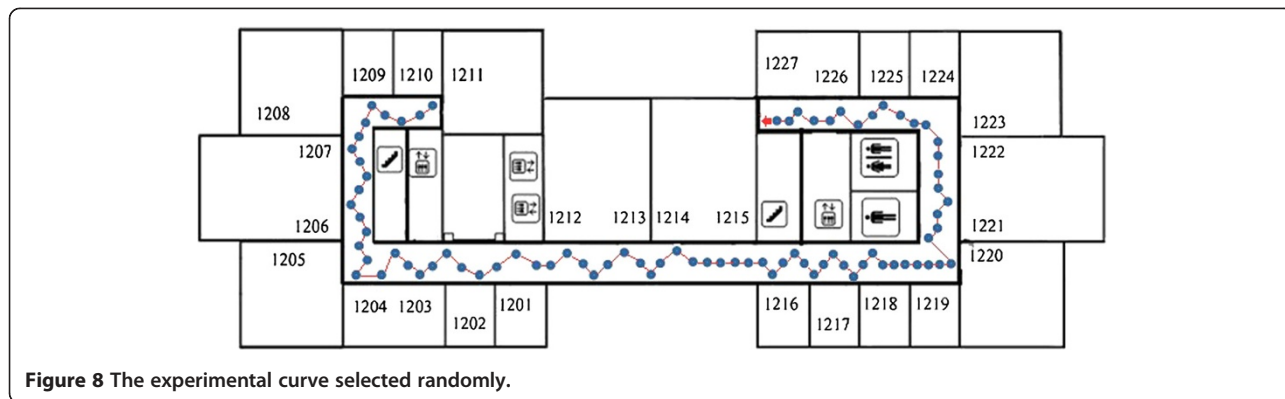**Figure 7 Results of affinity propagation clustering on RPs.**

Figure 8 The experimental curve selected randomly.

Last, by calculating Equation (13), one has

$$\mu^* = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{14}$$

$$\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{15}$$

The likelihood equations have a unique solution (μ^*, σ^{*2}) which should also be a local maximum point. This result can be interpreted that when $|\mu| \to \infty$ or $\sigma^2 \to 0$ or $\sigma^2 \to \infty$, the non-negative function $L(\mu, \sigma^2) \to 0$. Therefore, the maximum likelihood estimation of μ and σ^2 will be

$$\mu^* = \bar{X}, \tag{16}$$

$$\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \tag{17}$$

where X is defined as the set of RSS readings. Based on the statistical property of the maximum likelihood estimation μ^* and σ^{*2}, we can approximately recognize μ^* as the RSS fingerprint at each RP. With this idea, the mean of RSS reading e_i and the corresponding variance d_i from each hearable AP AP_i should be calculated for the construction of radio map in the off-line phase.

In the on-line phase, after collecting the new RSS readings {rss_i, $i = 1, 2, ..., L$}, according to Equation (18), we can calculate the probability of the RP (x, y) with respect to the i – th AP $P_i(x, y)$. In Equation (18), we have $\mu = e_i$ and $\sigma = d_i$.

$$P_i(x, y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(rss_i - \mu)^2}{2\sigma^2}}.(\sigma > 0) \tag{18}$$

Then, the probability of each RP $P(x, y)$ can be calculated. Finally, we will locate the user's position at the RP which has the maximum probability.

$$P(x, y) = \prod_{i=1}^{L} P_i(x, y) = \prod_{i=1}^{L}\left(\frac{1}{\sqrt{2\pi}d_i}e^{-\frac{(rss_i - e_i)^2}{2d_i^2}}\right) \tag{19}$$

5 Experimental results

Figure 3 shows the target indoor WLAN positioning environment for our testing. By using RSS readings collected from nine public APs (Cisco WRT54G), we will compare the performance of our proposed algorithm with other three typical positioning algorithm; (1) kNN positioning algorithm with K-means clustering (K-means + Knn); (2) probability distribution-aided positioning algorithm with K-means clustering (K-means + Probability Distribution); and (3) kNN positioning algorithm with affinity propagation clustering (Affinity Propagation + Knn). The dimensions of our testing area are 66×22 m^2.

5.1 Clustering results

In the experiments, we only focus on the situations that the damping factor is in the range of [0.5, 0.9] because only the damping factor falling into this range can guarantee that the affinity clustering results converge when the clustering process ended. Figures 4 and 5 show the variations of the numbers of iterations and clusters with respect to the values of damping factor respectively. To be clearer, we compare the numbers of iterations and clusters in the conditions of 9-AP (or all the APs), 5 APs (or about half the APs), and 3 APs (or the 3 APs which have the strongest RSS readings). As can be seen in Figures 4 and 5, the damping factor $\lambda = 0.65$ performs best in the computation cost (or with the smallest

Table 1 Results of statistical positioning errors

Algorithm	Mean (m)	Max (m)	Variance (m²)
K-means + Knn	3.38	6.72	2.82
K-means + Probability Distribution	3.11	6.70	3.24
Affinity Propagation + Knn	2.66	6.67	3.02
Proposed algorithm	2.23	6.52	2.52

Figure 9 CDFs of positioning errors.
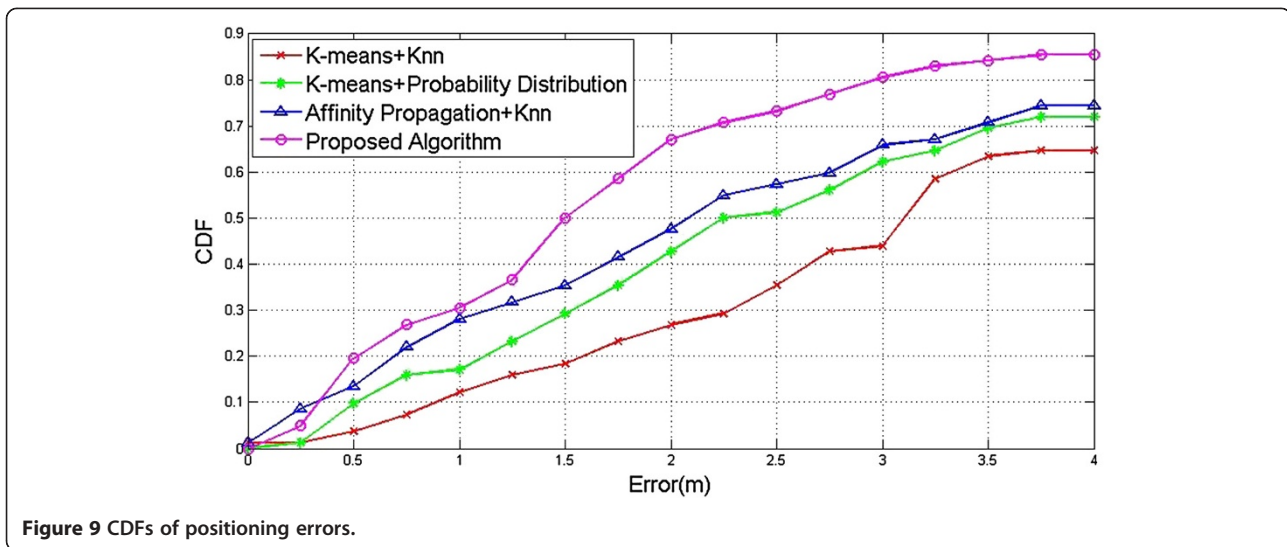
number of iterations), and meanwhile results in stable number of clusters.

Figure 6 shows the number of clusters with respect to the values of *preference* (or the parameter *p*) for the affinity propagation clustering. A large value *p* will result in the small number of clusters as expected. To distribute the RPs with the equal probability of being the cluster centers, we first set the value *p* as the median of the input similarities to generate a proper number of clusters, as described in Equation (4). After that, we can tune the values of *p* for the better clustering performance [21].

Figure 7 gives the results of the affinity propagation clustering on the RPs. The solid circles represent the calibrated RPs, and the RPs belonging to different clusters are labeled by different color. The 182 RPs have been clustered into seven clusters and the RPs in the same cluster are physically adjacent.

5.2 Positioning results
By randomly selecting 81 test positions in target area (see Figure 8), we can compare the error performance of the K-means + Knn, K-means + Probability Distribution, Affinity Propagation + Knn and our proposed algorithm in Table 1. Furthermore, the comparisons of the cumulative density functions (CDFs) of positioning errors are also illustrated in Figure 9. The red line in target area stands for the path that the user was walking.

5.3 Error analysis
Based on the previous experimental results which are conducted in a public indoor WLAN environment, we can observe that: (1) our proposed probability distribution-aided positioning algorithm has reduced the mean of errors by 34.02%, 28.3%, and 16.17%, respectively compared to the K-means + Knn, K-means + Probability

Distribution, and Affinity Propagation + Knn positioning algorithms; and (2) our proposed algorithm has also increased the confidence probability of errors within 3 m to 80.49% which is significantly larger than the probabilities 43.9%, 62.2%, and 65.85% achieved by the K-means + Knn, K-means + Probability Distribution, and Affinity Propagation + Knn positioning algorithms.

6 Conclusion
This paper proposes a new probability distribution-aided indoor positioning algorithm based on affinity propagation clustering. Compared with the conventional indoor fingerprint-based positioning algorithms, the positioning search space and computing cost are reduced. Because the affinity propagation clustering can be recognized as a preprocessing of the conventional fingerprint-based positioning algorithms, the proposed method can also be applied to the other fingerprint-based wireless positioning systems, like the RFID and mobile cellular network. However, due to the significant dependence on the RSS distributions and deployment of RPs, future research is required to optimize the layout of the APs and fingerprint modification for improving the positioning accuracy.

References

1. M Zhou, AK Wong, Z Tian, VY Zhang, X Yu, X Luo, Adaptive mobility mapping for people tracking using unlabelled Wi-Fi shotgun reads. IEEE Commun. Lett. **17**(1), 87–90 (2013)
2. S Gezici, A survey on wireless position estimation. Wirel. Pers. Commun. **44**(3), 263–282 (2008)
3. M Nicoli, S Gezici, Z Sahinoglu, H Wymeersch, Localization in mobile wireless and sensor networks. EURASIP J. Wirel. Commun. Netw. **197**, 2011 (2011)
4. M Zhou, Z Tian, K Xu, X Yu, H Wu, Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy. Expert. Syst. Appl. **40**(15), 6136–6149 (2013)
5. S Khodayari, M Maleki, E Hamedi, A RSS-based fingerprinting method for positioning based on historical data, in *Proc. IEEE SPECTS* (Ottawa, ON, 11–14 July 2010, 2010), pp. 306–310
6. A Kushki, KN Plataniotis, AN Venetsanopoulos, Kernel-based positioning in wireless local area networks. IEEE Trans. on Mobile Computing **6**(6), 689–705 (2007)
7. M Youssef, A Agrawala, The Horus WLAN location determination system, in *Mobile Systems, Applications and Services* (Seattle, WA, 06–08 June 2005), pp. 205–218
8. A Moragrega, P Closas, C Ibars, LACFA: an algorithm for localization aware cluster formation in wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **2011**, 121 (2011)
9. P Bahl, V Padmanabhan, RADAR: An in-building RF-based user location and tracking system. Proc. IEEE INFOCOM **2**, 775–784 (2000)
10. P Bahl, V Padmanabhan, A Balachandran, *Enhancements to the RADAR user location and tracking system*, 2000. Microsoft Research, no.MSR-TR-2000-12
11. M Youssef, A Agrawala, The Horus location determination system. Wirel. Netw. **14**(3), 357–374 (2008)
12. H Zang, F Baccelli, J Bolot, Bayesian inference for localization in cellular networks, in *Proc. IEEE INFOCOM* (San Diego, CA, 14–19 March 2010), pp. 1–9
13. M Zhou, Y Xu, L Tang, Multilayer ANN indoor location system with area division in WLAN environment. J. Syst. Eng. Eletron. **21**(5), 914–926 (2010)
14. X Ding, H Li, F Li, J Wu, A novel infrastructure WLAN locating method based on neural network, in *Proc. AINTEC* (Bangkok, 18–20 November 2008), pp. 47–55
15. L Mundo, R Ansay, C Festin, R Ocampo, A comparison of wireless fidelity (Wi-Fi) fingerprinting techniques, in *Proc. IEEE ICTC* (Seoul, 28–30 September 2011), pp. 20–25
16. A Teuber, B Eissfeller, T Pany, A two-stage fuzzy logic approach for wireless LAN indoor positioning, in *Proc. IEEE PLANS* (San Diego, CA, 25–27 April 2006), pp. 730–738
17. M Zhou, Y Xu, L Ma, Radio-map establishment based on fuzzy clustering for WLAN hybrid KNN/ANN indoor positioning. China Commun. **7**(3), 64–80 (2010)
18. R Battiti, T Nhat, A Villani, *Location-aware computing: a neural network model for determining location in wireless LANs* (Technical Report DIT-02-083, Ingegneria e Scienza dell' Informazione, University of Trento, 2002)
19. H Chan, M Luk, A Perrig, Using clustering information for sensor network localization, in *Proc. IEEE DCOSS* (Marina del Rey, CA, 30 June 30–01 July 2005), pp. 109–125
20. M Zhou, Y Xu, L Ma, S Tian, On the statistical errors of RADAR location sensor networks with built-in Wi-Fi Gaussian linear fingerprints. Sensors **12**(2), 3605–3626 (2012)
21. E Gokcay, J Principe, Information theoretic clustering. IEEE Trans Pattern Anal Mach Intell **24**(2), 158–172 (2002)
22. Y Wang, W Li, Y Sun, Wireless sensor network cluster locations: a probabilistic inference approach, in *Proc. IEEE ICAL* (Chongqing, 15–16 August 2011), pp. 76–80
23. BJ Frey, D Dueck, Clustering by passing messages between data points. Science **315**, 1 (2007)