Efficient Indoor Location Estimation using Multidimensional Indexes in Wireless Networks

Bong Gi Jun
Division of Computer and Information Engineering
Silla University, Busan, Korea

ABSTRACT

Since it is hard to use GPS for tracking mobile user in indoor environments, much research has focused on techniques using existing wireless local area network infrastructure. Signal strength received at a fixed location is not constant, so fingerprinting approach which use pattern matching is popular. But this approach has to pay additional costs to determine user location. This paper proposes a new approach to find user's location efficiently using an index scheme.

After analyzing characteristics of RF signals, the paper suggests the data processing method how the signal strength values for each of the access points are recorded in a radio map. To reduce computational cost during the location determination phase, multidimensional indexes for radio map with the important information which is the order of the strongest access points are used.

Keywords: Location based services, user location and tracking, fingerprinting.

1. INTRODUCTION

Location based services are booming and as marketing professionals to today's pervasive wireless technology[1], [2]. The GPS(Global Positioning System) does not work well for indoor application because satellite signals cannot penetrate buildings. The location of a mobile device can be estimated using radio signals transmitted or received by the device[3]-[7]. This paper focuses primarily on wireless local area networks.

For outdoor environments, the well-known approach to location estimation is based on angle and distance estimates from which a location estimate is deduced using standard geometry. However, the signal strength received at a fixed location in buildings is not a constant because the signal is affected by door, human body, etc.

Another type of location discovery is by pattern matching. Instead of estimating the distance between an access point and a mobile user, this approach tries to compare the received signal pattern against the training patterns in the database. Thus, this method is also known as the fingerprinting approach.

In the fingerprinting approach, a set of signal data of various locations is stored in a database. However, much training data is needed to improving the accuracy of location estimation. In location based services, moving objects[8] request their location information in real-time processing environment. Most of related works[3-6] have focused on accuracy of location estimation, and they used linear-time search algorithm.

Among many issues related to indoor location estimation, this paper focuses on the issues of reducing cost of searching the radio map, since this is one of the most important known

factors that determine the performance of the location determination algorithm. To address this issue, this paper proposes a new approach using the multidimensional index, which is based on the R-tree method. This paper defines master and slave axes to index sampling data which is the signal strength information from APs(Access Points). Three strongest APs are chosen as Master AP and Slave APs by the signal strength. The ordering of APs at observed locations is good information for determining user's locations.

The rest of the paper is organized as follows. Section 2 reviews approaches for indoor location estimation. Section 3 discusses the characteristics of the RF signal and costs of probability-based approaches. Section 4 presents a new approach using an index scheme for storing the collected signal sampling data at observation locations in off-line phase. Section 5 describes the query processing method for finding location of mobile user in real-time phase. Finally I summarize my contributions and give directions for future work in the section 6.

2. RELATED WORKS

The purpose of this section is to review the recent progress in indoor location estimation based on the fingerprinting approach. The main idea of the approach is to compare the received signals against those in the database and determine the likelihood that the device is currently located in a position. A typical solution has two phases :

. Off-line phase : The purpose of this phase is to collect signal data from each location and record them which are called as

"radio map"

. Real-time phase : With a well-trained positioning model, one can compare signal of unknown location with signal data to determine a user's location.

2.1 RADAR's approach

The simplest fingerprinting approach is the nearest neighbor in signal space (NNSS) approach[3]. In the first phase, only the average signal strength of each base station at each training location is recorded. As shown table 1, the mobile host records tuples of the form $(t, x, y, d, ss_i, snr_i)$ where i is the base stations during the off-line phase. They collected at least 20 signal strength samples from one location and computed mean value.

Table 1. Form of the table

Field	Meaning
x, y	user's coordinates
T	timestamp
D	direction of mobile device
SS	signal strength (dBm)
SNR	signal-to-noise (dB)

The problem is that collected value of set {signal strength, signal to noise} from same location is not always same many errors exist in the RF signal. Simple approach (mean value) cannot guarantee accuracy and single vector {ss, snr} per each location is not adequate.

2.2 Probabilistic approaches

Machine learning can be used in location estimation. A set of calibration data is collected from various locations, and is used in constructing a probabilistic model, which can be later used to estimate the probability distribution of the observation variable.

Roos[4] used Baye's theorem which is probability theory of conditional probability distribution for random variable. Roos[4] presented two ways (kernel method and histogram method) to implement the likelihood function. In the experiments performed, the two probabilistic methods produced slightly better results than the nearest neighbor method.

In the real-time phase, RADAR chose a simple linear-time search algorithm because their relatively a small data set and dimensionality. Related works[3, 4] focused on improving the accuracy of location estimation system.

Youssef[7] proposed a method by clustering training data in the database to reduce computational cost. But they didn't suggest how to use indexing or searching clusters as a set of locations sharing a common set of access points. The clustering techniques reduces the search space, however, if a coverage area is larger, then the computational cost increase linearly.

In the real field such as big buildings, hospitals or university campuses, lots of sampling data shall be stored to and retrieve from databases. To solve this problem, this paper was interested in efficient finding candidate set from large amount of sampling data in the indoor location-aware applications. This paper suggests an approach using a multi-dimensional index for

handling radio map efficiently.

3. PROBLEM DEFINITION

This section describes characteristics of RF signal which make the problem of WLAN location determination and high computational costs of probability-based approaches.

3.1 Characteristics of RF signal

A trend for the relation between distance and signal strength does exist. However, the ratio is unstable, because many uncontrollable environmental factors may incur significant errors. Figure 1 shows that the observed signal strength fluctuates over time.

Indoor radio propagation is difficult to predict because of the dense multipath environment and propagation effects such as reflection, diffraction, and scattering[9].
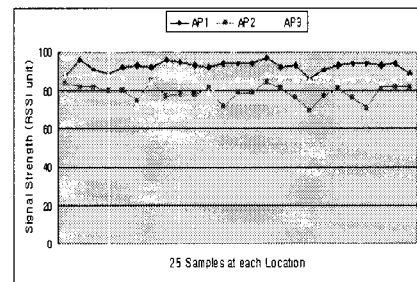

Fig. 1. Received signal strength from three access points

For this reason, pattern matching methods are applied to the problem of determining the location of a wireless device by measuring the signal strength values from a set of access points.

3.2 Complexity of probability-based approaches

The probability-based approach will try to maintain more complete information of signal strength distribution. By the Bayes rule, for computing location, calculating the probability over all target locations is too costly.

$$Cost = C_p \times N_{ap} \times N_l$$

where C_p is cost for computing probability in one probability function, N_{ap} is number of access points, and N_l is number of observed locations. For example, if N_{ap} is 10 and N_l is 100, every process of finding user location requires 1000 times of C_p. Related works used a simple linear-time search algorithm because their relatively small data set and focused the accuracy of location estimation.

4. MODELING OF RADIO MAP INDEX

After analyzing the received signal strength from the access points at selected locations, this section introduces how to handle the signal strength samples to be saved the radio map with a multidimensional index scheme.

Figures 2 and 3 show two different results for finding location where the distance of two locations is 3m. As shown figure 2, the received signal from AP2 can't be used to

determine location because average values and distributions are very alike. But the signal strength values from AP4 are usable for finding location.
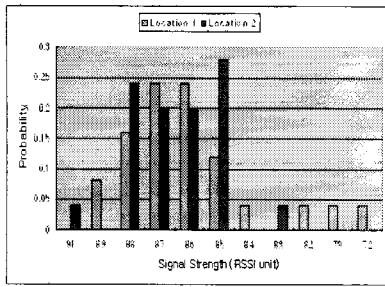

Fig. 2. Example of same signal pattern of an access point(AP2) at two different locations
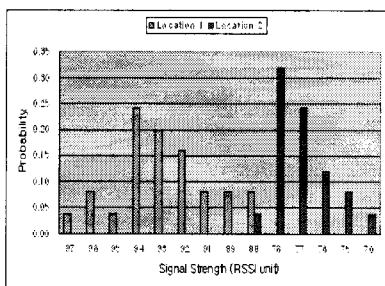

Fig. 3. Example of different signal pattern of an access point(AP4) at two different locations

To obtain a better accuracy, we need to handle signal strength values from *k* access points. The number of access points at a given location is varying with time.

The wireless LAN card received signal from sixteen APs(access points) on floor of the Computer Engineering Department building. Also the LAN card detected thirty six APs in my home. It is not good solution that received signals from all detected APs put it to a use of the location estimation.

Main idea is similar with the triangulation concept which considers signals from at least three beacons. Most of related works assign observation vectors, i.e., the signal strength from each access points to the likelihood function independently of each other.

This paper uses the master/slaves model with the signal strength values, and assumes that most of locations can receive signals from more than three APs. The master AP is the access point which sends the strongest average signal strength to the location. Then, the second strongest signal strength could decide to be the first slave AP. The received signal strengths from each one of APs are collected in off-line phase.
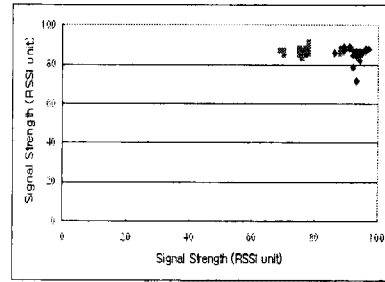

Fig. 4. A two-dimensional distribution chart of the signal strength values from two access points in different locations

Figure 4 shows the two-dimensional distribution chart that display signal strength values using in figures 2 and 3. As shown in figure 4, the signal strength values from APs can be made clusters and be distinguished in multi dimensions.

For obtaining good estimate of location, the signal data is stored from more than three access points. As dimension increases, the process of estimating location becomes more complex. Because my focus is low computational cost with the index scheme, the three strongest APs are used to determine user's location in my initial field test.

As shown table 2, set of APs{Master AP, Slave1 AP, Slave 2 AP} is different at each observed locations. In this work, I am mainly interested in the order of strongest APs for the location estimation problem. And the order helps to cluster the locations in three-dimensional domain.

Table 2. Example of three strongest APs and ASS(average signal strength) at three locations

Location	Master	ASS	Slave 1	ASS	Slave 2	ASS
1	AP1	91.64	AP2	88.60	AP3	84.76
2	AP4	92.60	AP2	85.60	AP5	79.00
3	AP6	92.92	AP7	92.48	AP8	88.92

In the figure 5, three points are coordinates of APs are made from tables 2, such as the coordinate of location1 is (AP1, AP2, AP3).
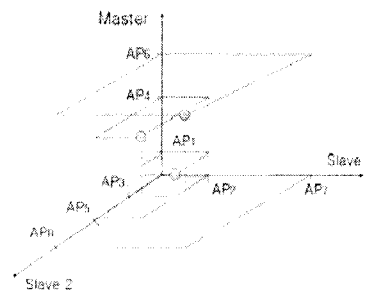

Fig. 5. Example of coordinates of APs in a three-dimensional index

If the order of strongest APs at a location is not changed, we can find user's location. But the strongest AP was changed in location3 sometimes, because observed signal strength fluctuates over time and difference between the received signal strength from two APs is small. The order changing is happened in 25% sampling data. In this case, coordinate (AP7,

AP6, AP8) is also saved to an index.

It is possible to be same that the order of strongest APs at the two location which are near. To solve this problem, the domain of axis is needed to have fine precision. RSSI values are combined with AP's ID. For example, the RSSI value(87) from AP(id is 5) is saved as 5.87(numeric value).

The domain of an index is not fixed because new APs can be installed in the building. R*-tree[10] was used for radio map. The MBB(Minimum Bounding Box)s enveloping RSSI values from APs at observation locations are stored in the leaf node.

5. QUERY PROCESSING FOR FINDING USER'S LOCATION

This section describes the query processing that makes candidate sets to estimate a device's location using the received signal strength values collected by the device.

In the real-time phase, a mobile user's location can be estimated given the signal strengths collected by the device from three strongest APs with a radio map using an index.

The estimation process has two step as follows :

① Filtering step : The purpose of this step is to determine a number of locations called candidate locations, each associated with a probability.

② Refinement step : Most probable location will be returned with the highest likelihood.

The focus of this paper is on indexing a radio map to reduce the computational requirements and increase the scalability of the system. This paper chose the multidimensional search algorithm using R-tree to consider the order of strongest APs.

Instead of using one signal strength sample collected by the mobile user's device from three strongest APs to estimate the user location, one can use a sequence of *n* samples at the same time. Since system has more information from the more samples, the accuracy should be enhanced.

Using observation sequences from three APs, the three dimensional MBB is made using minimum and maximum values of RSSI in each axis. The MBB is query's size for finding candidates of locations. If it can't find candidates, the MBB is extended 5 RSSI units to all directions, which is called the buffer query. Figure 6 shows a query's MBB and signal distribution collected at two locations in the filtering step.
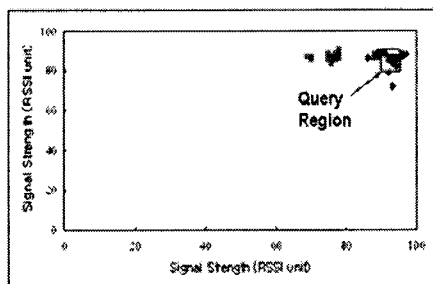

Fig. 6 Example of the query region in filtering step

If two more candidate locations are returned in the filtering step, probabilistic approach is used in the refinement step. It is

chosen the histogram method which is ease in implementation with low computational cost. Three histogram of the signal strength of three strongest APs are stored in a page, and the leaf node of R-tree has pointers of the page. The number of access page in the refinement step is equal to the number of candidate locations.

This paper didn't consider that the APs are out of order. If the location estimation system has the list of stable APs, this problem is solved simply.

6. CONCLUSION AND FUTURE WORKS

This paper has addressed characteristics of RF signal and showed relationships of sampling received signal from APs which are important factors for estimating user's location. This paper introduced two methods: an indexing method with the order of strongest APs in off-line phase and a location determination method with the buffer query in real-time phase. An approach using the multidimensional index would help improve its quick searching location in real world.

Nevertheless, initial field tests indicate that the suggested method is relatively robust with very low computational requirements, but I will develop further to reduce computational costs with the good accuracy.

REFERENCES

[1] J. Green, D. Betti, and J. Davison, *Mobile Location Services: Market Strategies*, Ovum Ltd, 2000.

[2] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," Computer, *Vol.34, No.8*, pp.57–66, 2001

[3] P. Bahl and V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," *IEEE INFOCOM, Vol.2*, pp.775–784, 2000.

[4] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation," *Int. J. of Wireless Information Networks, Vol.9, No.3*, 2002.

[5] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," *Int'l Conf. on Ubiquitous Computing (Ubicomp)*, pp.18-34, 2001.

[6] T. S. Rappaport, J. H. Reed, and B. D.Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Communications Magazine, Vol.34*, pp.33–41, 1996.

[7] M. Youssef, A. Agrawala, and U. Shankar. "WLAN location determination via clustering and probability distributions," *IEEE PerCom*, 2003.

[8] O. Wolfson, B. Xu, S. Chanmberlain, and L. Jiang, "Moving objects database: issues and solutions," *Proc. of Int'l Conf. on Scientific and Statistical Database Management*, pp.111-122, 1998.

[9] K. Pahlavan and P. Krishnamurthy, Principles of Wireless Networks: A Unified Approach, Prentice Hall PTR, Upper Saddle River, *New Jersey*, 2002.

[10] N. Beckmann and H. P. Kriegel, "The R*-tree: An efficient and robust access method for points and rectangles," *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pp.332-331, 1990.
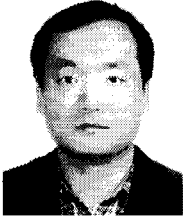
Bong Gi Jun
He received his B.S., M.S. and the Ph.D. degrees in Department of Computer Engineering from Pusan University, Busan, Korea in 1991, 1993 and 2003, respectively. From 1993 to 1998, he joined at Korea Telecom Research Center, where he worked as a manager of technical staff. In 2003, he joined the Division of Computer and Information Engineering, Silla University, Korea, where he is presently an assistant professor. His main research interests include Geographic Information Systems, Moving Object Databases and Tracking User location.