

WLAN Location Determination Using Probability Distributions with Search Area Reduction via Trilateration

David Sánchez⁽¹⁾, Miguel A. Quintana⁽¹⁾, Juan Luis Navarro⁽²⁾

⁽¹⁾Department of Ingeniería Telemática
⁽²⁾Department of Señales y Comunicaciones
University of Las Palmas de Gran Canaria
Spain
{dsanchez,mquintana}@dit.ulpgc.es, jnavarro@dsc.ulpgc.es

Abstract—Location methods of users in WLAN are being studied because of they can provide information to other applications to taking better decisions and improving their performance. Location methods should be designed to obtain high accuracy and low computational cost. However, it is not an easy task to achieve due to high accuracy requires many calculations. In this paper, we present a location system for IEEE 802.11 infrastructure networks based on probability distributions that uses trilateration technique to determine a starting point of search. Once the starting point is inferred the search area will be limited the closest locations to this point.

Keywords—WLAN location, received signal strength, probability distributions, trilateration

I. INTRODUCTION

Nowadays, the development in mobile computing devices and the advances in wireless technologies allow using portable computers in many situations of the daily life. One of the open challenges of mobile computing is the current location estimation of users [1]. In this respect the location-based computing has made possible applications with the capability to sense their location and modify their setting and functions accordingly [2]. A location tool can be used to develop applications related to navigation, security, health and entertainment.

During the last years many authors have dedicated their efforts to solve the problem of location determination in IEEE 802.11 networks. In this context indoor location is still considered a fundamental problem in mobile computing for which the high accuracy requirement is not clearly resolved [3].

Several types of sensing modalities have been used for location determination, such as angle of arrival (AOA), time difference of arrival (TDOA) and received signal strength (RSS) of the RF waves. The latter is one of the most used modalities due to its availability in many devices thus making unnecessary extra sensors. The RSS measurements is one of the most used information sources in location computation. Techniques like trilateration, weighted

centroids, etc. use this information. More sophisticated methods use model-based approaches. Among them fingerprinting is one of the most promising for indoor localization. This is mainly due to the dense multipath effect, the propagation effects such as reflection, diffraction and scattering and non line-of-sight environment that cause inaccurate angle and time estimation [4]. Besides, fingerprinting technique does not need additional resources because of network interfaces provide information about received signal strength from each access point. For that reason, most of research works use a model based on signal strength. This is usually done in two phases. In the first phase, normally termed *offline phase*, a database of the received signal strength in each location is built. And in the second one (*online phase*), given a sample of RSS and a classification system which uses database information, the location is determined. Most of works differ in the latter phase.

In [5], a data classification method based on support vector machine (SVM) is used with acceptable performance. Its main drawback is that it requires an extra time for training SVM. In [6], two methods for location determination in indoor and outdoor environments are used: training data method based on nearest neighbour algorithm and a propagation method based on RSS decreases logarithmically with distance. The authors indicate that the first method produces more consistent results in the outdoor environment. However, its accuracy can suffer if there are numerous points with similar training data. This may be the situation in indoor environments due to multipath. The propagation model is more accurate in an open line of sight environment but less accurate when sources of interferences were introduced.

Probabilistic methods are also used to location determination. In [7], authors use probability distributions combined with ray-tracing approach to account for absorption and reflection characteristics of various obstacles. The main advantage of this method is the use of only an access point. However, the authors recognize that

the maximum accuracy is achieved in a given room inside a building. In [8], probability distributions are also used, with high accuracy, taking account the high correlation between samples from the same access point with clustering techniques to reduce the computational requirements. Depending on the number of access points and the location of them, the size of clusters can be large such that many possible locations may be inside, and therefore, it can be heavy computationally. It has specially to be considered if it is designed for handheld devices where the processor power is not high and the energy availability is limited. In order to reduce the search area and therefore, the number of possible locations to explore, in this paper we propose to use the trilateration technique to determine the starting point of search. In this way, the search area is limited to the locations closest to this point. Therefore, in this paper we present a location determination method based on probability distributions that uses trilateration technique to reduce the search area.

The rest of the paper is organized as follows. In Section 2, we describe the trilateration technique used for determining the starting point of search. Next, in Section 3, we explain the probability distributions method used to infer location of device. In Section 4, we show some experimental results. Finally, we sum up the conclusions and we present the future work.

II. TRILATERATION TECHNIQUE

Trilateration is a technique by which the location of a device can be determined by measuring the radial distance of the received signal from two or three different points. Trilateration is sometimes used in cellular communications to pinpoint the geographic position of a user. In IEEE 802.11 infrastructure networks the access points are usually in a fixed position allowing communicating devices in a coverage area. If the position of three access points (vertices

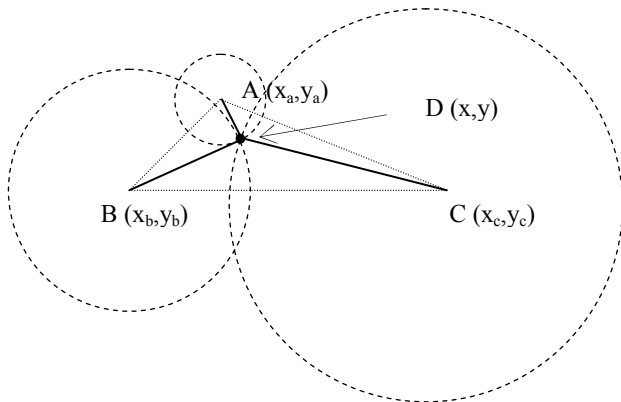


Figure 1. Trilateration method

A, B, C of the triangle in Fig. 1) and the distances $\overline{DA}, \overline{DB}$ and \overline{DC} are known, then a trilateration method can be used to infer D position. This is done by finding the intersection of the three circumferences which centres are A, B and C vertices. Therefore, given coordinates of each access point (x_i, y_i) and distances from the portable device to each of them (d_i) , the position of a device can be obtained by the nonlinear system indicated in (1).

$$(x - x_i)^2 + (y - y_i)^2 = d_i^2 \quad (1)$$

$$i = a, b, \dots, n$$

In order to obtain the position of a portable device it is necessary to know the location of at least three access points. If two access points are known then the system returns two possible solutions, and if only one access point is available, the possible positions are located in the circumference with radius equal to the access point distance.

To solve (1) is necessary to determine the distance from portable device to each access point. For doing that, we are based on the logarithmic path loss (transmitted power minus received power) which is given by the conventional distance power loss [9] as:

$$P = S + 10 \times n \times \log(d) + \sum L_w \quad (2)$$

where P is the path loss in dB from access point to device, S is the path loss in dB to 1 metre from access point, n is the attenuation factor which depends on the propagation environment, d is the distance in metres between transmitter and receiver and accounts for the L_w is the penetration loss when there are walls between transmitter and receiver. Therefore, given a RSS value, we need to know the attenuation factor and the path loss to determinate the d distance using (3) which is obtained from (2). These values may be estimated in an empirical way for each access point because of they depend on the propagation environment and usually transmitters with different characteristics are used and located in different places.

$$d = \log^{-1} \left(\frac{P - S - \sum L_w}{10 \times n} \right) \quad (3)$$

To be able to obtain the value of n many samples of RSS have to be collected at each possible location for each transmitting point and when there is only direct vision among transmitter and receiver. Next, from measurements and applying (4) we estimate an attenuation factor for each access point and for each location. Finally, an average equivalent attenuation factor is calculated by each access

point.

$$n_{di} = \left(\frac{P - S}{10 \times \log(d_i)} \right) \quad (4)$$

On the other hand, if in some locations there is not direct vision among transmitter and receiver, the penetration loss has to be estimated. They are obtained by the difference between the RSS when the transmitter is in direct vision and the RSS when there are one or several walls among them. Once we have all parameters and a given set of RSS values collected in a location from k access points the distance from the device to each access point is estimated using (3). Next, knowing the position of the access points and the distances from portable device to them, (1) can be solved.

III. LOCATION DETERMINATION USING PROBABILITY

In this section we describe the method based on probability distributions to infer location. We describe the process performed in the offline and online phases, and the benefits of using trilateration to reduce the searching area.

A. Offline training phase

During the first phase a database of the received signal strength (RSS) in each location and for each access point is built. It is normally called radio map. To perform this task, first, we have to define the possible location points in the environment, and then, a process of RSS collection must be carried out in each of these points. For each training position it is stored a set of n samples (r_1, \dots, r_k) received from k access points (AP_1, \dots, AP_k) where r_i is a signal strength value from AP_i .

Based on these values and for a given location l the probability $P(r_i/l)$ can be estimated using the normalized histogram of the access point AP_i at this location. Probability values for each access point and location are

stored in radio map. Fig. 2 and Fig. 3 show an example of the signal strength normalized and collected to ten and twenty metres from an access point, respectively. As it can be seen, the signal propagation varies with the distance. Several authors [8][10] indicate that distribution of RSS can be closely a lognormal distributions. However, we have noticed that it depends on the environment and how the data collecting is performed. For example, in our experiments we take samples in four different directions in each location. That is, data are gathered when the portable device is oriented to north, east, south and west. Fig. 4 and Fig. 5 show an example of 200 data collected at each location where every fifty samples the portable device is turned ninety degrees. As it can be seen, RSS average value changes in each direction and in each location. Therefore, we can say that distribution of RSS can be closely a mix of lognormal distributions (Fig. 2 and Fig. 3), and small spatial variations affect to signal distribution.

The size of set of samples for each location and how they are collected can influence on the accuracy of the system. Therefore, in order to obtain good estimate of probability distribution it is necessary a large set of samples for each position collected in different directions. This phase requires a lot of time to collect received signal strength.

B. Online location determination phase

In this phase the location determination is calculated from a given set of RSS values obtained from some access points. The location with higher probability is returned as the user or device position. For that, Baye's theorem is used to estimate the observed values in each position.

Given a sample of RSS values $R=(r_1, \dots, r_k)$ gathered from k access points we want to find a location l such that $P(l/R)$ is maximum. Applying Baye's theorem as in [8] $P(l/R)$ can be formulated as:

$$P\left(\frac{l}{R}\right) = \frac{P\left(\frac{R}{l}\right) \times P(l)}{P(R)} \quad (5)$$

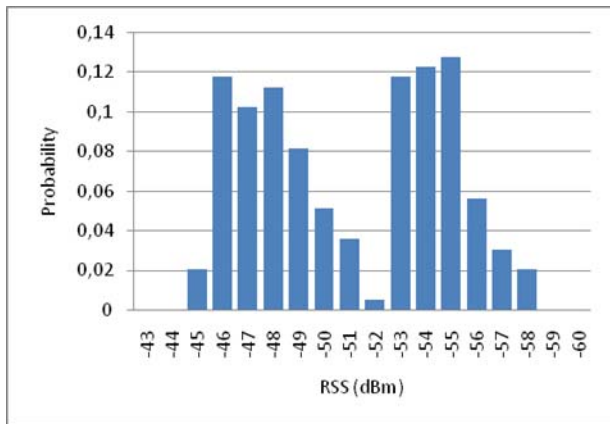


Figure 2. A sample histogram to 10 metres from access point.

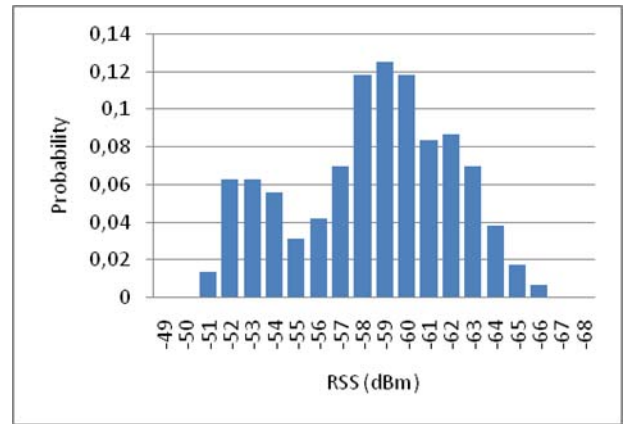


Figure 3. A sample histogram to 20 metres from access point.

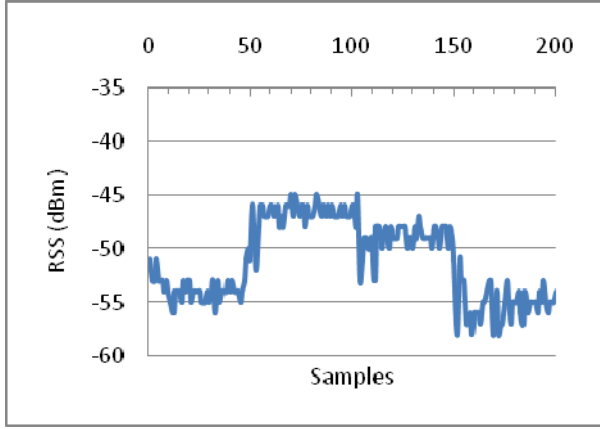


Figure 4. RSS collection with different orientations to 10 metres from access point.

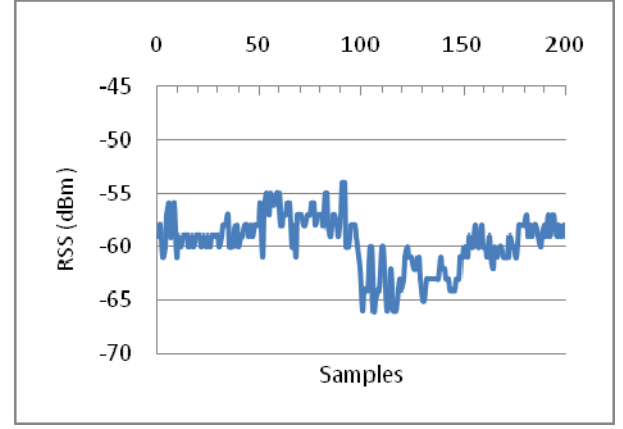


Figure 5. RSS collection with different orientations to 20 metres from access point.

Since $P(R)$ has the same value for all locations, then the above equation can be simplified as:

$$P\left(\frac{l}{R}\right) = P\left(\frac{R}{l}\right) \times P(l) \quad (6)$$

On the other hand, the probability that a user is a location $P(l)$ is equal for all locations. Therefore, this factor influences in the same way in all $P(l/R)$. Then, it can be defined as:

$$P\left(\frac{l}{R}\right) = P\left(\frac{R}{l}\right) \quad (7)$$

If we assume that access points do not run on overlapped channels, the RSS values received from them can be considered as independent. Therefore, $P(R/l)$ can be calculated as follows:

$$P\left(\frac{R}{l}\right) = P\left(\frac{r_1}{l}\right) \times \dots \times P\left(\frac{r_k}{l}\right) = \prod_{i=1}^k P\left(\frac{r_i}{l}\right) \quad (8)$$

where $P(r_i/l)$ was calculated from samples collected in each location in the offline training phase.

C. Search Area Reduction

As it was mentioned in previous section, location determination based on probability distributions is estimated in the location where $P(R/l)$ is maximized. Therefore, this value has to be calculated via (8) for all possible locations that are covered by access points. The number of possible locations can be large and it depends on the environment, the position of access points and the area size of possible locations. For example, if in a determined location is collected RSS values from one only access point, the search space is extended to the coverage area of that access point. However, if beacons packets are received from two or more access points the search space is reduced to the intersection

of all coverage areas. Fig. 6 shows an example of the intersection of coverage area of two access points. In order to carry out the search in the intersection area a previous classification has to be done in offline training phase to distinguish possible areas of search given a set of RSS values. In any case, it implies a computational cost and therefore power consumption that has to be taken into account if location determination is estimated in mobile devices.

In order to reduce the number of operations we propose to use the trilateration method such that the intersection of the three circumferences will give us the starting point to begin the search. Due to the equivalent attenuation factor used for solving (3) is calculated by approximation from empirical data and the environment conditions may change in a random way the results obtained in (1) may differ to the correct solution. Therefore, the search will be done in a circle of radius s which centre is the solution of (1). The value of radius s is a tradeoff between the search area and computational cost. If RSS values are collected from two access points the intersection of the two circumferences will give us two starting points to search. In Fig. 7 the searching areas are shown as red zones whose centers are the intersections of circumferences with radius equal to the distances from access points to device. Finally and in the worst case, if an access point is only available the search has to be performed along to the ring formed on having displaced the circle of radius s in the circumference of radius equal to the distance from that access point to device.

Table 1 shows the search area reduction versus number of access points, where r is the radius of the coverage area of an access point, a is the area of a possible location, d is the distance from the access point to device that collects RSS values, and s is the radius of a circle where the search is done (red circle in figure 7). The value of radius s is a tradeoff between the search area and computational cost, but it is always smaller than the radius r of coverage area; the ratio used is greater than ten. The value of distance d depends on

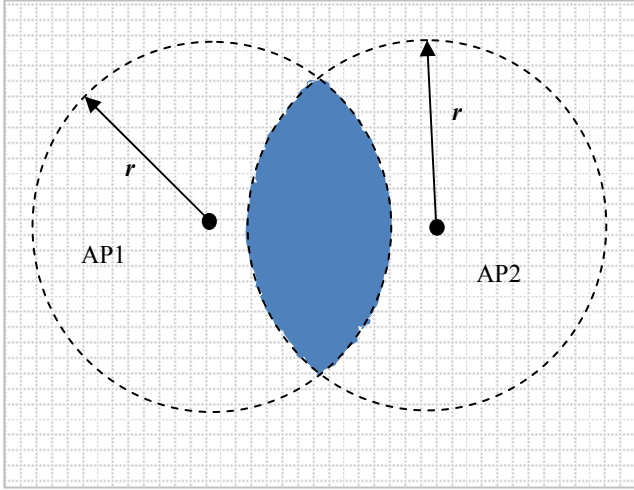


Figure 6. Intersection of coverage area.

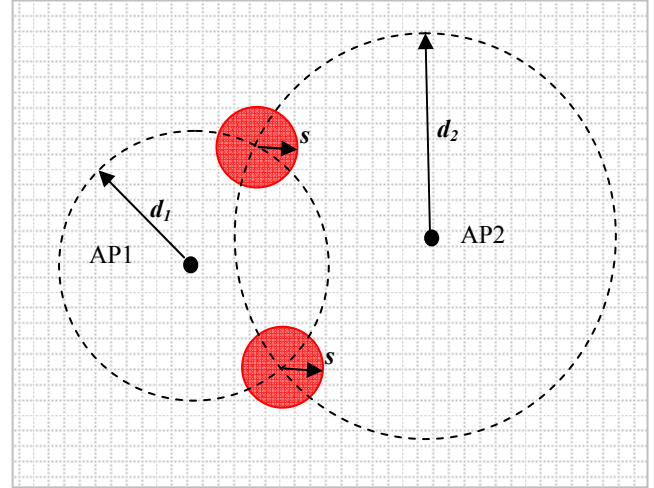


Figure 7. Search space reduction.

the location of the user, but it will always be less or equal to r .

TABLE I. SEARCH AREA REDUCTION

Available Access Points	Normal	Trilateration
1	$\pi r^2 / a$	$4 \pi d s / a$
2	$(2\pi/3 - \sqrt{3}/2) r^2 / a$	$2 \pi s^2 / a$
3	$(AP1 \cap AP2 \cap AP3) / a$	$\pi s^2 / a$

IV. RESULTS

We performed our experiments in the second floor of the Telematic Engineering Department at the University of Las Palmas de Gran Canaria. The floor has a dimension of 63 metres by 10 metres. The layout of the floor and the position of three access point (indicated with a triangle) are shown in Fig. 8. Our experiments were focused only in the corridor which is approximately 2 metres wide.

Before testing our system, we have built a radio map. For that, we have defined 126 cell locations of approximately 1 m^2 in the corridor. In each cell we took about 200 RSS samples from each access point where every 50 samples the portable device was turned ninety degrees. Measurements

were made at different hours during several days and when there was only direct vision between transmitter and receiver. From these data, a histogram is obtained for each cell and each access point. From the average values of RSS in each cell, knowing the path loss to 1 metre and the transmitted power for each access point we calculated the equivalent attenuation factor in each cell using (4). Next, we estimated with these values the average attenuation factor for each access point. Low standard deviation was obtained. The values obtained are shown in the table 2.

TABLE II. ATTENUATION FACTOR FOR EACH ACCESS POINT

WireLess_GAC_A	1.65
Telematica	1.53
WaveLAN	1.88

Due to the access points are situated in the corridor the coverage area of them covers all it, that is, in each cell we have samples of three transmitters. Therefore, the penetration loss because of the walls was not estimated and they were not used in (3). Results were focused to evaluate the accuracy and computational cost.



Figure 8. Layout of the testbed floor

A. Accuracy

In order to evaluate the accuracy of the location system in the corridor, we put the portable device at different locations. At each position and at different time, the method was invoked several times whereas portable device remained static. Table 3 shows the average success percent. As it can be seen, the location method has at least a 49% of probabilities to return the correct position, and it is closer to 90% accuracy to within three metres.

TABLE III. AVERAGE SUCCESS PERCENT

Location (m)	0 m	1 m	2 m	3 m	$\geq 4m$
5	71%	12%	8%	1%	8%
15	54%	23%	7%	4%	12%
25	55%	19%	13%	3%	10%
35	51%	15%	11%	7%	16%
45	59%	15%	13%	4%	9%
55	49%	19%	17%	5%	10%

B. Computational Cost

The computational cost of the system depends on the starting point estimated by the trilateration method and the size of radius s of the circle with centre in this point. From table 3, we can deduce that the optimum value of s is 3 metres because of it represents a 90% of accuracy. We evaluated the computational cost as the number of cells to explore in order to find a location where the probability is maximized. In our testbed floor, a value of s equal to 3 metres represents fourteen cells.

In the worst case and if trilateration method was not used 126 cells must be to explore for finding the cell where the probability is maximized. Therefore, location system presented in this paper has a computational cost of 11% regarding the worst case, that is, a reduction of 89%. Some authors [8] use clustering to classify different searching areas. In our case, since beacons packets are received from three access points in all cells, we only can to differentiate three search areas. Each area is identified by the nearest access point. If we suppose three similar areas, each zone would have approximately the same number of cells, 42. Therefore, the method presented in this paper would have a computational cost of 31% regarding cluster, that is, a reduction of 69%.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a location system in IEEE 802.11 networks using probability distributions with an acceptable accuracy, 90% within three metres. The best feature of this work is the low computational cost due to the search area reduction. This low computational cost is obtained via trilateration technique, and its value is at least 69%. Therefore, the location system could be implemented in handheld devices with energy constraints.

Since the range and the measurement of RSS depend on the network card, in our experiments we have used the same wireless card for creating the radio map and testing the method. For that reason and due to manufactures use different cards for each other, in our ongoing work we are planning to design a new technique based on the difference between RSS levels to be used with different network cards. Besides, we are also studying the possibility to implement Kalman filters to minimize spatial and temporal variations of RSS.

ACKNOWLEDGMENT

This work has been partially supported by University of Las Palmas de Gran Canaria under Contract ULP GC 07/030, by European Sixth Framework Programme under Contract FP6-2005-IST-5-034642 and Department of Ingeniería Telemática.

REFERENCES

- [1] A. Zomaya, "Mobile Computing: Opportunities for Parallel Algorithms Research", 15th IEEE IPDPS, pp. 144–147, 2002.
- [2] R. Want and B. Schilit, "Expanding the Horizons of Location-Aware Computing", IEEE Computer, pp. 31–34, August 2001.
- [3] L. Zong, D. Kotz, R. Jain and X. He, "Evaluating next-cell predictors with extensive Wi-Fi mobility data", IEEE Transactions on Mobile Computing, vol 5, 12, pp. 1633–1649, December 2006.
- [4] P. Bahl and V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system", IEEE Conference on Computer Communications (INFOCOM'00), pp. 775–784, March 2000.
- [5] C. Wu, L. Fu, F. Lian, "WLAN Location Determination in e-Home via Support Vector Classification", IEEE Conference on Networking, Sensing & Control, Taiwan, pp. 1026–1031, 2004.
- [6] M. Emery and M. Denko, "IEEE 802.11 WLAN based real-time location tracking in indoor and outdoor environments". Canadian Conference on Electrical and Computer Engineering, pp: 1062–1065, 2007.
- [7] G. V. Zaruba, M. Huber, F. A. Kamangar and I. Chlamtac "Indoor location tracking using RSSI readings from a single Wi-Fi access point". Wireless Network, Springer, vol. 13, pp: 221–235, 2007.
- [8] M. Youssef and A. Agrawala, "The Horus WLAN location determination system", IEEE International Conference on Mobile Systems, Applications, and Services, 2005.
- [9] A. Santamaría, F. López-Hernández, *Wireless LAN Systems*, Artech House, 1994.
- [10] K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location determination", International Symposium on Wireless Pervasive Computing, 2006.