Modelling An Enhanced Routing Protocol For Wireless Sensor Networks Using Implicit Clustering Technique

Idigo Victor[1], , Azubogu A.C.O[2], Oguejiofor Obinna3,Nnebe Scholar[4]

[1]Department of Electronics and Computer Engineering, Nnamdi Azikiwe University,Awka
Anambra State(234),Nigeria

[2]Department of Electronics and Computer Engineering, Nnamdi Azikiwe University,Awka
Anambra State(234),Nigeria

[3]Department of Electronics and Computer Engineering, Nnamdi Azikiwe University,Awka
Anambra State(234),Nigeria

Abstract: The localization of sensor nodes can be a very enabling technology that can help in improving the performance of many algorithms designed for wireless sensor networks (WSNs). This work is geared towards developing a positioning system that uses received signal strength based on fingerprinting technique. The proposed system models the signal strength distribution received from the sensor nodes using non-parametric or Gaussian distribution. The probabilistic Bayesian technique was employed as the localization algorithm for the basic model. The result obtained shows that there is an improved median error, appropriately 1.0 meter, compared to 2.5 meters for the nearest neighbour (NN) algorithm. Implicit clustering technique was used to enhanced the result obtain from the basic model. The performance of the basic model was enhanced by more than 18% for the static model and more than 10% for the mobile model. Finally using the enhanced model reduces the average number of operations per location estimate by more than 30%.

Keywords: Sensor nodes, Bayesian Technique, Clustering, Location estimate, Error probability

1.0. Introduction

Wireless sensor networks(WSN) are based on network of devices that can be densely deployed in an aggressive and inaccessible environments to sense the environment, and monitor with high accuracy the physical phenomena[1]. Each one of these devices is called a sensor node. They have limited processing speed, storage capacity and communication bandwidth.

In many WSN scenarios, the random deployment of hundreds of sensor nodes without localization hardware raises the problem of determining the topology of the network in terms of the outer boundary and the boundaries of the communication nodes [2]. Existing boundary recognition algorithms are able to determine these boundaries with certain accuracy. However, they only work for extremely dense networks and involve high computational and message complexities,

In multihop wireless networks, it is energy efficient to choose long paths along a series of short hops rather than short paths along a series of long hops. However, even though efficiency is always of paramount interest, it is not the only one [3]. Communications performance is also very important. By choosing many short hops we may lower the energy expenditure, but only to a certain degree; since delay increases, processing energy increases and control overhead increases. Therefore, the choice of how to incorporate energy is not as clear as it seems [3,4].

A useful distinction refers to whether energy is treated as a cost function or as a hard constraint [3], in the former case, the objective of the designer is to minimize the amount of energy per communication task, treating energy as an expensive but in exhaustive resource. However when energy is a hard constraint, the designer sees energy as a limited resource that will be exhausted[3,5]. In this case, the task is more complicated since there is a need to satisfy conflicting objectives; maximizing the longevity of the network versus communication performance such as throughput, total data delivery, etc.

The localization of sensor nodes can be a very enabling technology and can provide help to improve performance of many of the algorithms designed for WSNs. For example, in geographic routing protocols, the location information (in the form of coordinates) is used to select the next forwarding host among the senders neighbours. In rescue applications, rescue personnel can perform their task only if location of the hazardous event (reported by nodes) is known. Some related work in this area include the following: In [4 , 5] some methods for estimating unknown node positions using exclusively connectivity induced constraints are presented. These methods are only suitable for location determination with beacons. Some works reports are about an ingenious algorithm based on GPS free positioning. This algorithm explores only each node's knowledge of the neighbours and produces a coordinate system for each node and for the network. One major drawback of this approach is that the nodes do not know the physical direction of the coordinate system. In [6] the authors address the deployment by aircrafts of node groups and determine the positions of a node through the neighbours considering that the node is located close to the drop place of the node groups more represented. The bigger drawback is that the deployment does not always act like a model[4,6]. The authors use a mobile beacon to scan through the network, broadcasting its position while it passes. Although that is a good idea, it is not always possible to move beacons around a deployment area. [7] Addresses the problem of simultaneous localization, tracking and calibrations using probabilistic Bayesian filtering. This was reported to be a very good algorithm for ultrasound localization, but still lacks accuracy when using radio connections. However, this technique had been employed with some success in the field of robot localization [8]. In practice, the Bayesian localizer proves more accurate that the deterministic techniques such as the nearest neighbour(NN) algorithm because it takes into account more information from the training data during the data collection phase and filters the output using motion model.

This work employs the probabilistic Bayesian approach for the basic localization algorithm. The result obtained from the basic model was enhanced using the implicit clustering technique. Clustering of radio map locations was introduced as an approach to reduce the computational requirements of the location determination algorithm, improve accuracy and achieve scalability. The results show that using clustering reduces the average number of operations per location estimate by more than an order of magnitude.

2. Experimental Set up and Methodology

2.1 Experimental test bed

The test bed is located at the first floor of the 3-storey Administrative building of Nnamdi Azikiwe University, Awka. The floor has a dimension of 20m by 18m in an area of 360 sqm and segmented by a square of 1x1 meters as shown in fig 1. The deployment of the sensor nodes are shown marked AP1, ... AP4 in figure 1.

The transmitter and receiver were placed in different positions with respect to each other in the test bed. We used MSP430 mote which is developed by crossbow for the equipments. The mote employs the CC2420 ; which is a single-chip 2.4 GHz IEEE 802.15.4 RF transceiver with DSSS baseband modem of 2Mchips/s and 250kbps effective data rate with digital Received Signal Strength Indicator (RSSI), Link Quality Indicator (LQI) and MAC support. The transmitter nominal output was set to 0dBm and the receiver sensitivity was set at -90dBm. CC2420 has a built-in Received signal strength indication providing a digital value that can be read from the eight bit, signed 2's complement RSSI. RSSI_VAL register. The RSSI value is always averaged over eight symbol periods (128µs) in accordance with [9].The RSSI register value RSSI.RSSI_VAL can be referred to the power P at the RF pins by using the following equations;

P= RSSI_VAL + RSSI_OFFSET (1)

RSSI OFFSET is found empirically during system development and is approximately -45. For example, if reading a value of -20 from the RSSI register , the RF input power is approximately -65dBm.

The link quality Indication (LQI) measurement is a characterization of the strength and/or quality of a received packet as defined by [9]. Using the RSSI value directly to calculate the LQI value has the disadvantage that for example a narrowband interferer inside the channel bandwidth will increase the LQI value although it actually reduces the true link quality. CC2420 therefore also provides an average correlation value for each incoming packet. Software must convert the correlation value to the range 0-255 defined by [9], e.g. by calculating: LQI= (CORR-a).b . The Variables a and b are found empirically based on PER measurement as a function of correlation value.

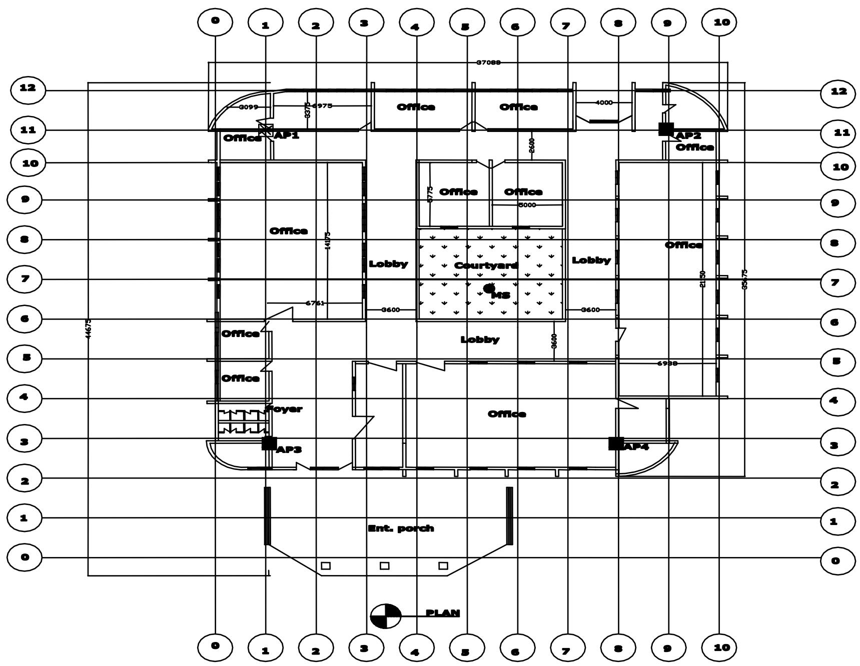

Figure 1: Experimental Test bed

2.0 Methodology

In the experiment , four different data packet sizes(20 bytes,30 bytes,50 bytes and 70 bytes) were transmitted over the wireless link of interest once every 100ms with transmitter-receiver separations of one meter. Ten measurements were taken for every T-R separation(fig 1). Ten such measurements were taken for each of the packet size but with 1-meter increment of the transmitter-receiver separation, up to 10 meters. In addition we sampled the Received signal strength indicator(RSSI) for every byte of data received. Averaging these values over an entire packet, an estimate of averaged received signal power for a packet was calculated.

This work is geared towards developing a positioning system that uses received signal strength based on finger printing method which is dependent on building a database. The reason for developing such a model is that today there is no way to develop and evaluate the performance of a positioning system except running massive measurements. Accuracy of the system is closely related to the number of nodes in the database and the distribution of them. In using fingerprinting method, it is required that a grid-network be built prior to any location estimation. After building the database for a new location , the new metric is measured irrespective of the viewed location and compared it with the database to find the best node, which could be referred to as desired point.

The proposed system models the signal strength distribution received from the sensor nodes using Gaussian distribution. The main advantage of the Gaussian technique is the efficiency of calculating the location estimate. [4,7,10] showed analytically that this technique is optimal among all discrete- space radio map-based location determination systems. The probabilistic Bayesian approach was employed as the localization algorithm for the basic model. The results obtained from the basic model is enhanced using the implicit clustering technique[11,12].

3. Problem Formulation and Implementation of Basic Model

Two vectors are normally used in estimating the location of the mobile station (MS). The first vector consists of samples of the RSSI measured at the mobile stations from N sensor nodes in the area. This Vector is denoted as $S = [s_1, s_2, s_3 ..... s_n]$. The indoor positioning system estimates the mobile's location using the sample RSSI vector.

The second vector that forms the finger printing of the location, consists of the true means of all received signal strength random variable at a particular location from the N sensor nodes and recorded in the location database. We call it the location fingerprint or the average RSSI vector and denoted by $R = [r_1, r_2, r_3, ...r_n]$.

Let X be a 2 or 3 dimensional physical space. At each location xεX we can get the signal strength from N sensor nodes. The problem of the basic model becomes, given a signal strength vector $S(x) = [s_1, s_2, s_3 ..... s_n]$. We want to find the location xεX that maximizes the probability $P(x/S)$.

To solve this problem , the probabilistic method of finger printing such as the Bayesian approach to WLAN localization was used.

This has been employed with some success in the field of robot localization. If I_t is the location at time t, 0_t is an observation made at t (the instantaneous signal strength values) and N, the normalization factor that ensures all probabilities sum to 1, then for localization , Bayes rule for static situation can be written as:

$$P\left(\frac{I_t}{0_t}\right) = P\left(\frac{0_t}{I_t}\right). P(I_t). N \qquad (2)$$

Equation (2) implies that the probability of being at location I given observation 0 is equal to the probability of observing 0 at location I , and being at location I in the first place. During localization, this conditional probability of being at location I is calculated for all finger prints. The most likely location is then the localizer's output. In order to calculate $P\left(\frac{I_t}{0_t}\right)$ in equation (2), it is necessary to calculate the two probabilities on the right hand side of the equation. $P\left(\frac{0_t}{I_t}\right)$ is known in Bayesian terms as the likelihood function. This can be calculated using the signal strength map. For each fingerprint, the frequency of each signal strength value is used to generate a probability distribution as a likelihood function.

Markov localization suggests using the transitional probability between locations. This probability is described as:

$$P(I_t) = P(\frac{I_t}{I_{t-1}})\ P(I_{t-1}) \qquad (3)$$

In other words, $P(I_t)$ is the sum of the transitional probability from all locations at t-1 to I at current time t , multiplied by the probability of being at these locations at t-1. $P(I_{t-1})$ is known from previous localization attempts. We calculate $P(\frac{I_t}{I_{t-1}})$ using a motion model, for instance, for a walking person the simplest and effective approach is to calculate the probability based on how far the user can move between t and t-1 [9]. The result obtained using Bayesian approach was compared to the use of nearest neighbour technique as reported in RADAR, an in-building RF based user location and tracking system[12].when 20 byte of data was transmitted figures (2) and (3) show the performance results using Bayesian and NN techniques for static and mobile localization, respectively. Static localization is performed for targets not expected to move and takes the prior probability as the uniform distribution. For mobile localization, the prior probability was calculated using a simple motion which caused the accuracy to be significantly improved compared to the NN approach . There is an improved media error when summarizing RSSI information as Gaussian approximately1.0 meter, compared to 2.5 meters for the NN in this test. In addition to improve accuracy the Bayesian localizer provide a frame work for the integration of other sensor nodes, infer red or mobile phone signal strength, can be integrated into the model by running the same Bayesian update equation on a shared probability distribution

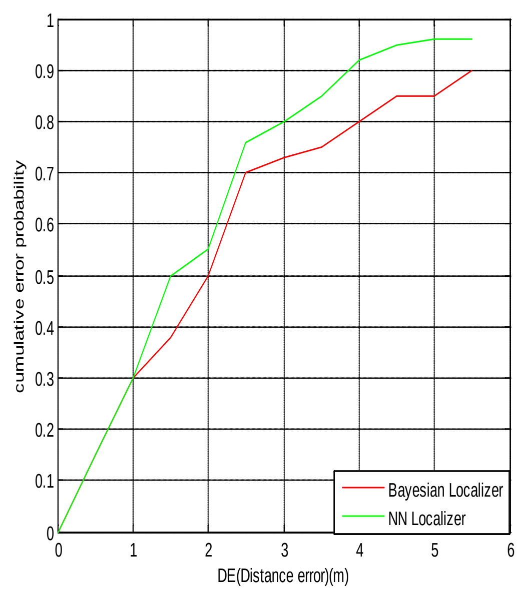

Fig 2: Cumulative Error probability for static localization

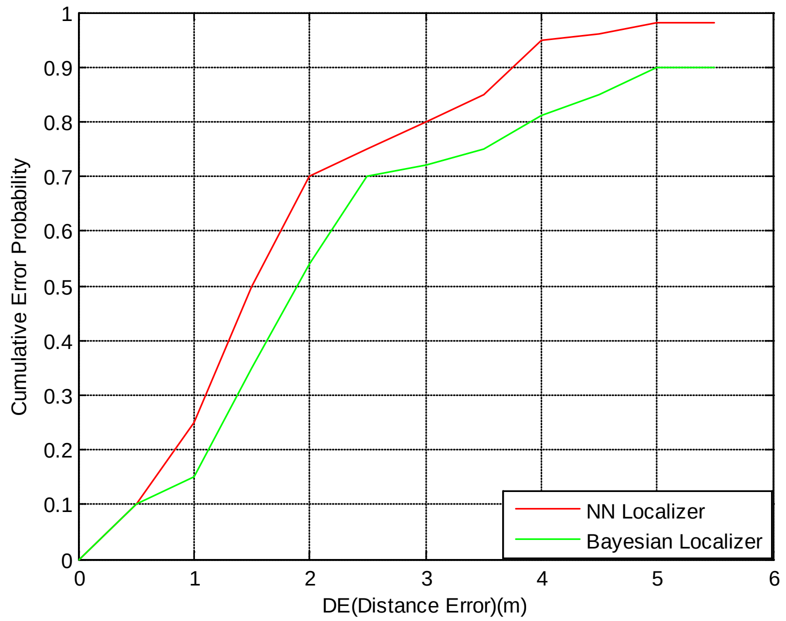

Fig 3: cumulative error probability for mobile localization

Figure 4 shows that the probability of error versus distance error for the four different data sizes. The results shows that the probability of error value is a function of the number of bytes transmitted.

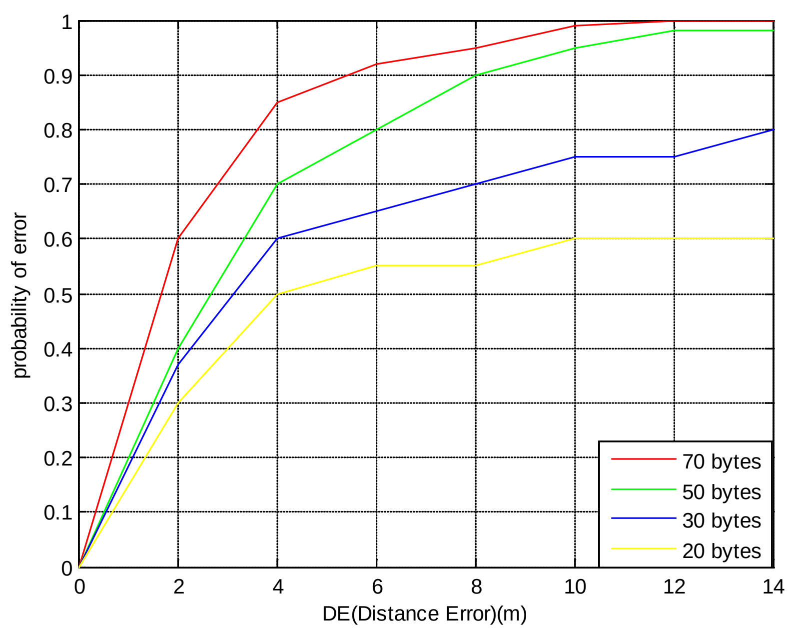

Figure 4: Comparative analysis of the probability of error versus Distance error for different data sizes

4. The Enhanced Model

Radio map location clustering was used in this work for the enhancement of the performance results defined by the basic model. This technique reduces the computational requirements of a WLAN location determination algorithm[5].

A cluster is defined as a set of locations sharing a common set of access points. This common set of access points is called the cluster key. The problem can be stated as : Given a location x, we want to determine the cluster if which x belong. Two clustering approaches are presented in [5]: explicit clustering where the system must determine the clusters during the offline or data collection phase as a separate step, and ,the implicit clustering where no special processing is performed in the offline phase but rather during the location determination phase, the system performs clustering implicitly. The explicit clustering technique was reported in [5,8] as producing slightly better accuracy than that of the implicit clustering technique. However, the average number of operations performed per location estimate for the clustering technique is much lower than the corresponding number of the explicit clustering technique [8].

In order to conserve energy during the location determination process, this work adopted the implicit clustering approach as the technique used in the enhanced model. The enhanced algorithm works as follows: considering a sequence of RSSI values from each sensor node , we start by sorting the sensor nodes in descending order according to their average RSSI values. Then for the node with the strongest average RSSI value , we calculate the probability of each location in the radio map set given observed RSSI sequence from this node alone. This gives us a set of candidate locations (location that have non-zero probability). If the probability of the most probable location is "significantly" higher (according to a threshold) than the probability of the second most probable location, we return the most probable location as the location estimate, after consulting only one node. If this is not the case, we go to the next node in the sorted sensor node list. For this node ,we repeat the same process again, but only for the set of candidate location obtained from the first sensor node. Finally, the algorithm returns the most probable location in the candidate list that remains after consulting all the nodes.

Figures (5) and (6) gives a comparative analysis of the cumulative error probability for the Nearest neighbour (NN), the basic model and the enhanced model algorithms. Figure (5) is the analysis for the static case while figure (6) for the mobile case.

Figure (7) shows the variation of the average number of operation per location estimate for the basic and enhanced models

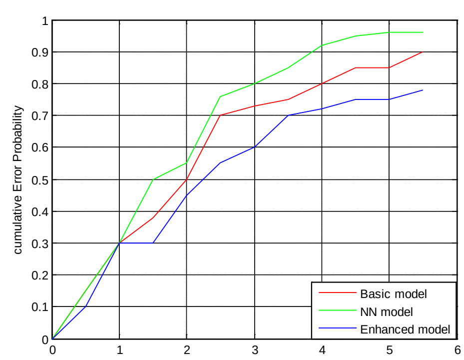

Figure 5: Cumulative Error Probability Analysis for NN, Basic and Enhanced models (Static Localization)

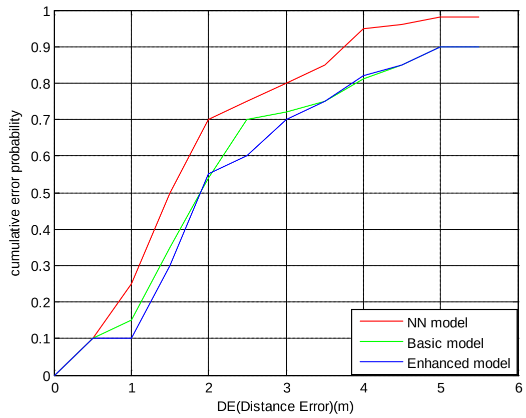

Figure 6: Cumulative Error Probability for NN, Basic, and Enhanced models(Mobile Localization)

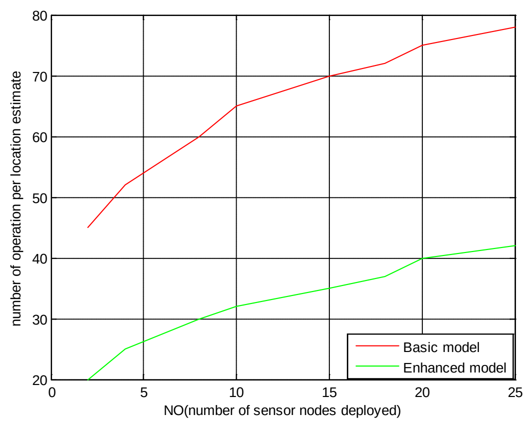

Figure 7: Comparative Analysis of number of operations required per location estimate versus the number of sensor nodes deployed

5. Conclusion

This work presents an empirical modelling of an enhanced indoor positioning system that uses RSSI/fingerprinting technique. The system models the signal strength distribution received from sensor nodes using non- parametric or Gaussian distribution. The main advantage of the Gaussian technique is the efficiency of calculating the location estimate.

The probabilistic Bayesian algorithm was employed for the basic model. The result obtained was compared to the NN algorithm used in RADAR. This is an improved median error when summarizing RSSI information Gaussian appropriately 1.0 meter, compared to 2.5 meters for the NN in this test. The performance analysis of the proposed model with respect to the data size transmitted was shown using probability of error technique, which is the probability that the location technique would give an incorrect estimate. Result shows that the performance of the system slightly decreases with increase in data size. The performance of the basic system was enhanced by more than 18% for the static model and more than 10% for the mobile model using the implicit clustering technique. Results also show that using the enhanced model reduces the average number of operations per location estimate by more than 30%.

References

[1]. Ming Zhang, Yanhang LU and Chenghong Gong, (2008), “Energy Efficient Routing protocol based on clustering and least square tree in Wireless sensor Networks”. International Conference on computer science and software Engineering pp 361 – 364.

[2]. Bager Zarei , Mohammed Zeynali and Vahid Majid Nezhad, (2010), IJCSI vol. 7, issue 4. www.ijcsi.org

[3]. Akyildiz I.F., weilians.,(2002), “A survey on Sensor Networks” IEEE communications Magazine, 40(8): 102-114.

[4]. Elgamal A.et al(2004), “A framework for monitoring Bridge and Civil

infrastructure" proceeding of 3rd China-Japan-US symposium structural health and control.

[5]. Cesare Alippi, Giuseppe Anastasi, Mario Di Francesco, Manuel Roveri (2009), " Energy Management in Wireless Sensor Networks with Energy-hungry Sensors". IEEE instrumentation and Measurement Magazine. Vol.12 no.2.

[6]. David Culler, Deborah Estrin, Mani Srivastavai(2004), "Overview of Sensor Networks", Special issue in sensor networks, IEEE computer 37(8), pp 41-49.

[7] Jan Rabacy, et al(2007), " Pico radio: communication/ computation piconodes for sensor networks". Technical Report, Electronic Research Laboratory, pp 7-22.

[8]. Kamik A. And Kumar A.(2004),"Iterative Localization in wireless adhoc sensor network : one dimension Case", in proceedings of the international conference on signal processing and communications.

[9]. IEEE std. 802.15.4 (2003): Wireless Medium Access Control(MAC) and Physical Layer(PHY) specifications for low rate wireless personal Area networks(LR-WPANs).

[10]. Monstafa Abdel A.V. (2004), "HORUS : A WLAN- Based Indoor location determining system", PhD dissertation, Worchester, polytechnic Institute.

[11]. Sichitiu M.L, Ramadurai V.(2003), "Localization of Wireless sensor networks with a mobile beacon: a mobile beacon based Bayesian approach to localizing network nodes.

[12]. Bahl P. And Padmanabhan V.(2000), "RADAR: An in-building RF-based user location and tracking system, proceedings of IEEE infocom, Tel-Aviv, Israel. Vol 2 pp 755-784.