

LOCATOR - Location Estimation System For Wireless LANs

Ankur Agiwal^{*}
ankur@iitd.ernet.in

Parakram Khandpur[†]
parakram@iitd.ernet.in

Huzur Saran
saran@cse.iitd.ernet.in

Department of Computer Science & Engineering
Indian Institute of Technology
New Delhi, India

ABSTRACT

With the wide spread growth of mobile computing devices and local area wireless networks, wireless network providers have started to target the users with value-added services based on the users' location information. Thus, location awareness and user tracking in indoor wireless networks has become an increasingly important issue. To address the issue, we have developed LOCATOR, a radio frequency based system for location estimation of users in indoor wireless networks. Our system works by building a radio map of the network site, which involves taking signal strength samples at various points in the wireless network, and then using this radio map, the system estimates the user's current location from the value of his current observed signal strength. Although the use of an RF-based technique for location estimation is not new, but LOCATOR is unique in the way it builds and manages the radio map to process location queries. Moreover, in addition to using ideas of probability to model the problem, we have devised a multi-level clustering based algorithm that work in conjunction with an interpolation scheme for a more efficient and accurate location estimation. We have tested our system on 802.11b wireless network testbeds and have been able to achieve an accuracy of location estimation to within 4 feet of the actual user location with 90% probability.

Categories and Subject Descriptors:

C.2.1 [COMPUTER-COMMUNICATION NETWORKS]:
Network Operations – *Network monitoring, Network management*

General Terms: Algorithms, Management, Measurement, Performance, Experimentation

^{*}Now a graduate student at the Department of Computer Science, UNC, Chapel Hill

[†]Now a graduate student at the Department of Computer Science, Stanford University

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WMASH'04, October 1, 2004, Philadelphia, Pennsylvania, USA.
Copyright 2004 ACM 1-58113-877-6/04/0010 ...\$5.00.

Keywords: location-aware services, user location, clustering, location management, wireless LAN, interpolation

1. INTRODUCTION

The widespread deployment of wireless local area networks and the increasing popularity of light-weight mobile computing devices has lead to an increased interest in location-aware applications and services. The goal of these applications and services is to enable the user to interact more effectively with his environment. Examples of such value added service include, but are not limited to, services like displaying the map of the immediate surroundings and guiding a user inside a building. A great deal of research has, thus, focused on developing services and architectures for location-aware systems and on the problem of location estimation and user tracking in indoor wireless environments.

The techniques used for location estimation include GPS [1], infrared[3] based systems, ultrasonic[7] based systems, wide-area cellular[2] based systems and radio frequency (RF) [9, 11, 10, 12] based systems. Among these the RF based technique for location estimation and user tracking has gained a lot of attention due to its good accuracy and ease of deployment without the need for any specialized additional hardware.

Radio frequency (RF) based location estimation systems work in primarily two phases: radio map building phase and location estimation phase. In the radio map building phase, a large number of signal strength samples from the base stations are collected at various locations in the wireless network to build a radio map database, which is essentially a mapping from the physical space to the signal space. In the location estimation phase, the system uses a user's currently observed signal strength and the radio map to estimate his location.

In this paper, we present LOCATOR: an RF-based system for locating and tracking users in an 802.11 wireless LAN framework. However, the core idea of LOCATOR is generic enough to be used for other wireless network technologies also. The key ideas that make LOCATOR accurate and scalable are: (a) Use of clustering to fragment the radio map to improve scalability and reduce the computational cost for location estimation (b) Use of probability distributions and location-based clustering along with interpolation scheme to enhance the accuracy

We have evaluated our system on multiple 802.11b testbeds but in this paper, due to space constraints, we present the results for only one of the testbeds. The experimental results of LOCATOR are very encouraging and provide estimate of user location to within 4 feet with 90% probability.

The remainder of this paper is organized as follows. Section II briefly reports related work on location estimation and user tracking. Section III describes the methodology of LOCATOR. In section IV, the evaluation results of LOCATOR on our testbeds are discussed. Finally, section V concludes the paper and gives directions for future work.

2. RELATED WORK

Many systems in the recent past have attempted to tackle the problem of location estimation and user tracking in wireless networks. In this section, we briefly discuss some of the major approaches.

The Active Badge System [3], an infrared-based system, was an early and significant contribution to the field of location-aware systems. In this system, a badge worn by a person emits a unique infrared signal every 10 seconds. Sensors placed at known positions within a building pick up the unique identifiers and relay the location manager software. While this system provides accurate location information, it suffers from several drawbacks: (a) it scales poorly due to limited range of infrared and (b) it incurs significant installation and maintenance costs.

Another system based on infrared technology is described in [4]. In this system, infrared transmitters are attached to the ceiling at known positions in the building. An optical sensor on a head-mounted unit senses the infrared beacons, which enables the system software to determine the user's location. This system suffers from similar drawbacks as the Active Badge System.

There have also been several location estimation systems for wide-area cellular based networks [2]. The technological alternatives for location cellular phone users involve measuring the signal attenuation, the angle of arrival (AOA) and/or the time difference of arrival (TDOA). While these systems have been found to be promising in outdoor environments, their effectiveness in indoor environments is limited by the multiple reflections suffered by the RF signal, and the inability of off-the-shelf and inexpensive hardware to provide fine-grain time synchronization.

Systems based on the Global Positioning System (GPS) [1] have also been proposed and have been found to perform extremely well in outdoor environments. Unfortunately, buildings block GPS signals and therefore, GPS does not operate in indoor environments.

A class of location estimation systems that have recently become very popular is the radio frequency (RF) based system. These systems basically work in two phases: (a) the radio map building phase and (b) location estimation phase. These phases were briefly discussed in section 1. The main advantages of these systems are their ease of deployment without the need of any specialized hardware and their reasonably good accuracy.

RADAR [9], developed at Microsoft Research, was the first system to use the RF-based technique for location estimation and user tracking in wireless LANs. RADAR builds a single monolithic radio map for the wireless network site and carries out a k -nearest-neighbor (k typically being 3 or 4) search in the signal space. It estimates the user's loca-

tion as the mean of the physical locations of the k neighbors. This approach does not give very high accuracy and moreover, since RADAR searches the entire radio map, it has the drawback of having a very high computation cost for user location estimation.

Horus [12], another RF-based location estimation system, improves upon the performance of RADAR by using their Joint Clustering technique and clustering locations in the radio map based on the strongest signal, in case there are more than 3 base stations in range. By the use of these techniques Horus achieves a significant performance gain, both in terms of accuracy as well as computation cost, over RADAR. The main drawback of the Horus system is that its location estimation algorithm can estimate a user's location only to a location that was initially used while building the radio map. Unlike LOCATOR, which incorporates interpolation techniques for location estimation, Horus system's performance degrades significantly if the granularity of the radio map is increased.

3. LOCATOR

Our location estimation system, LOCATOR, works in two phases. In the first phase, it builds a radio map of the site where the wireless network is deployed. The radio map is a mapping between the physical locations and the observed signal strength values at these locations from various base stations. In the second phase, LOCATOR uses the radio map to estimate the location of a user, given its current observed signal strength values from various base stations. These two phases are described in greater detail in the following subsections.

3.1 Radio Map Building Phase

In this phase, a database is built that maps the physical space to the signal space. The database is built only for a few selected locations, which is then used by LOCATOR for location estimation of users as described in section 3.2.

In the following subsection, the signal strength sampling strategy is discussed, followed by the explanation of the methodology adopted for fragmenting the radio map database into clusters to reduce the computation cost while answering location search queries.

3.1.1 Sampling strategy

The IEEE 802.11b standard [6] uses radio frequencies in the license-free 2.4 GHz band in which other devices like microwave ovens, cordless phones and Bluetooth devices also operate and can, thus, be a source of interference [5]. Moreover, the phenomenon of multi-path fading also causes temporal variations in the observed signal strength. Due to multi-path, the transmitted RF signal reaching the receiver through different paths has different phase and amplitude. These different components combine and produce a distorted version of the transmitted signal. Other factors causing the temporal variations include movement of people, closing and opening of doors and other environmental changes. Due to all these reasons, the wireless channel has a highly unpredictable nature.

On carrying out an experiment to study the temporal variations in the signal strength samples collected at a fixed location from a base station, variations upto 14 dBm were observed as depicted in figure 1.

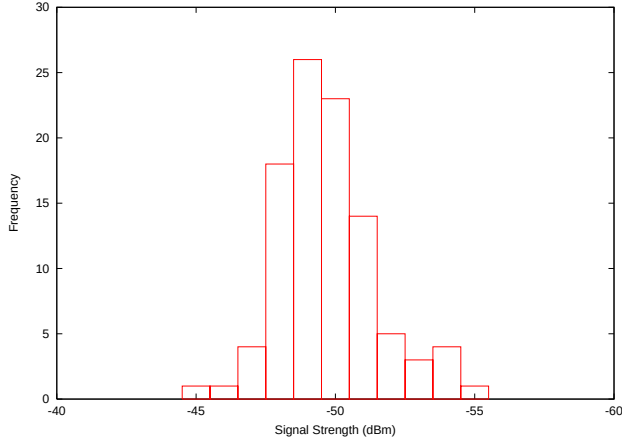


Figure 1: Temporal variation of signal strength

Many earlier RF-based approaches like RADAR [9] have used only the mean signal strength values for location estimation. But due to the highly unpredictable nature of the wireless channel, using only the mean value of signal strength for location estimation is not sufficient for obtaining accurate results. Therefore, instead of taking only the mean signal strength value for building the radio map, a signal strength probability distribution function for various locations due to each of the base stations is defined as suggested in [12]. Assuming that the base stations operate on non-overlapping wireless channels, the signal strength probability distribution functions of the base stations at a particular location can be assumed to be independent of each other.

For simplicity only 3 base stations are assumed to be present in the network, though the model can be extended to any number of base stations. Thus, the probability of the signal strength at a location (x,y) being (q_1, q_2, q_3) due to the base stations can be expressed as:

$$P_{x,y}(q_1, q_2, q_3) = P_{x,y}^1(s = q_1) \cdot P_{x,y}^2(s = q_2) \cdot P_{x,y}^3(s = q_3)$$

where q_i is the signal strength value and $P_{x,y}^i(s)$ is the signal strength probability distribution function at location (x,y) corresponding to the i^{th} base station.

To obtain $P_{x,y}^i(s)$, a frequency distribution table corresponding to location (x,y) is constructed, as shown in Table 1. The frequency distribution table is built by taking a large number of signal strength samples at location (x,y) . For instance, $P_{x,y}^i(s = S_k^i)$ would be given by:

$$P_{x,y}^i(s = S_k^i) = \frac{f_k^i}{\sum_{j=1}^n f_j^i} \quad (1)$$

3.1.2 Clustering of radio map

The radio map database can be of a very large size depending on the size of the area where the wireless network is deployed and the granularity of the radio map database. By granularity of the radio map database, we refer to the physical closeness between two consecutive points where the signal strength samples are taken while building the radio map database. Therefore, to make the technique scalable to larger areas, the database is fragmented into location-based clusters.

AP 1 (dBm)	freq.	AP 2 (dBm)	freq.	AP 3 (dBm)	freq.
S_1^1	f_1^1	S_1^2	f_1^2	S_1^3	f_1^3
S_2^1	f_2^1	S_2^2	f_2^2	S_2^3	f_2^3
S_3^1	f_3^1	S_3^2	f_3^2	S_3^3	f_3^3
S_4^1	f_4^1	S_4^2	f_4^2	S_4^3	f_4^3
...	...	...	...	...	...
...	...	...	...	...	...

Table 1: A sample frequency table for location (x,y)

Since location estimation and user tracking requires the users to periodically collect the signal strength samples from the various base stations in range and query the radio map to estimate its location, therefore, an upper bound can be assumed on the maximum distance a user can move between two such consecutive queries. This upper bound depends on the time period between two consecutive queries and the mobility pattern of the user. Therefore, to process users' queries for location estimation, instead of searching the complete radio map database, only the portion corresponding to the neighborhood of the user's previous location in the physical space needs to be searched. This strategy leads to significant decrease in the computation cost incurred for each query, leading to improved performance.

Depending on the upper bound on the maximum distance a user can move between two consecutive queries, the radio map database is fragmented into location-based clusters as shown in figure 2. Thus, if a user is in cluster (i,j) as per his previous query's estimate, then for estimating his current location the search is limited to cluster (i,j) and its neighboring clusters.

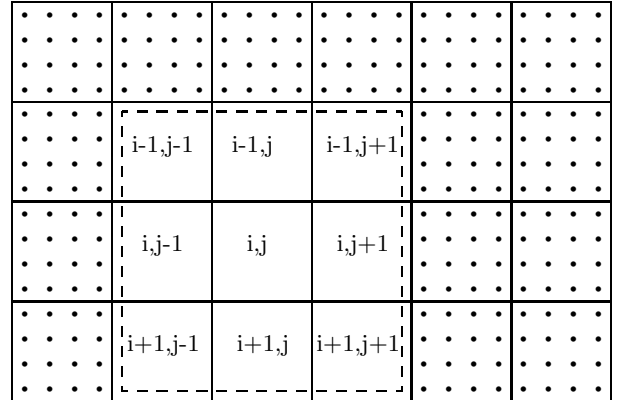


Figure 2: Fragmenting the radio map into location-based clusters

Let the upper bound on the maximum distance a user can move between two consecutive queries be d . For LOCATOR, the value of d , which is a tunable parameter, has been taken to be 15 meters. This value is based on two factors: (a) the time period between consecutive queries, which has been set to 3 seconds and (b) the maximum speed of a user, which has been taken as 5 m/s based on a survey conducted by us on the mobility behavior of users inside buildings.

Thus, with the use of location-based clustering in the radio map building phase, the computation cost greatly reduces and makes LOCATOR scalable to larger wireless networks and also allows a finer granularity of the radio map. Till now, an assumption that there are only 3 wireless base stations in the wireless network has been made. In case there are more than 3 base stations then a first level clustering based on strongest signal base station can be done as described in [12]. After the first level clustering, location-based clustering can be used without any modification.

3.2 Location Estimation Phase

In the location estimation phase, the users periodically query the radio map database with their observed signal strengths (q_1, q_2, q_3) from the 3 base stations. To estimate a user's location, LOCATOR works as follows:

1. Performs a k -best search in the 3-dimensional signal space using the previous estimate of the user's location and (q_1, q_2, q_3)
2. Clusters the k points based on their physical closeness
3. For each cluster, estimates the most probable location of the user and its likelihood
4. Chooses the location with the highest likelihood as the current estimate of the user's location

The above steps are described below in greater detail.

3.2.1 K -best search

To reduce the computation cost, the k -best search is performed only on a portion of the radio map database. If the user's previously estimated location lies in some cluster (i, j) then the search space comprises of the cluster (i, j) and its neighboring clusters only as depicted in figure 2. For k -best search, a k -nearest neighbor search in the signal space is performed using the average values of the signal strengths stored in the radio map. Two points are considered to be the closest in the signal space if their Euclidean distance is the least i.e. (q_1, q_2, q_3) and (s_1, s_2, s_3) are the closest in signal space if the $(q_1 - s_1)^2 + (q_2 - s_2)^2 + (q_3 - s_3)^2$ is minimum. Algorithms exist that perform k -best search in $O(k \log n)$ time where n is the size of the sample to be searched.

3.2.2 Clustering based on physical closeness

The k candidate points $\{v_i\}_{i=1}^k$ obtained from the k -best search described in subsection 3.2.1, are now grouped into clusters based on their physical closeness.

Definition 1. The *diameter* of a cluster is the maximum distance between any two nodes in the cluster.

Clustering of nodes is done in a manner that each cluster has its diameter less than the *maximum diameter*, D . The optimum value of D is obtained empirically as discussed in section 4.4.2. Thus, if $d(u, v)$ denotes the distance between nodes u and v , then the clustering algorithm partitions the k points into m disjoint clusters $\{C_i\}_{i=1}^m$, such that

$$d(u, v) \leq D \quad \forall u, v \in C_i, 1 \leq i \leq m \quad (2)$$

Definition 2. The distance between two clusters C_i and C_j is denoted by $d(C_i, C_j)$ and is defined as:

$$d(C_i, C_j) = \max \{d(u, v) : u \in C_i, v \in C_j\} \quad (3)$$

Claim 1. If diameter of each of C_i and C_j is less than D and $d(C_i, C_j) < D$ then diameter of $C_i \cup C_j$ is less than D .

PROOF. Consider any edge $(u, v) \in C_i \cup C_j$

Case I: u and v belongs to same cluster. Say, $u \in C_i, v \in C_i$ then by definition of diameter $d(u, v) < D$.

Case II: u and v belongs to different clusters. Say, $u \in C_i, v \in C_j$ then by definition of distance $d(u, v) < D$. $\square$

Algorithm for clustering:

```

1   $C_i = \{v_i\}$  for all  $1 \leq i \leq k$ 
2   $S = \{C_1, C_2, \dots, C_k\}$ 
3   $i = k$ 
4  while  $(i > 1)$  do
5      begin
6          find  $i, j$  s.t.  $\Delta = \min\{d(C_i, C_j)\} \forall C_i, C_j \in S$ 
7          if  $(\Delta < D)$ 
8              begin
9                   $C_{new} \leftarrow C_i \cup C_j$ 
10                  $S = S - \{C_i\} - \{C_j\}$ 
11                  $d(C_{new}, C) = \max\{d(C_i, C), d(C_j, C)\} \forall C \in S$ 
12                  $S = S + \{C_{new}\}$ 
13             end
14         else stop
15          $i \leftarrow i - 1$ 
16     end
```

The above clustering algorithm terminates either when all the clusters have been merged into one single cluster or when no two clusters exist such that the distance between them is less than the *maximum diameter* D .

Steps 1, 2 and 3 take constant time and the *while* loop in step 4 can loop atmost k times. Step 5 is $O(k^2)$ operation while steps 6, 8, 10, 11 and 12 are constant time operations. Steps 7 and 9 each take $O(k)$ time. Thus, the overall time complexity of the clustering algorithm is $O(k^3)$. But, k is small so number of operations are less.

3.2.3 Estimation of most probable location and its likelihood for each cluster

For each cluster, the most probable location and the likelihood of the user being at that location is calculated. For a cluster consisting of n points, Lagrange multipliers l_{1i} , l_{2i} and l_{3i} corresponding to each of the n points are calculated as described below, where (q_1, q_2, q_3) is the signal strength tuple that the user has provided in his query request.

$$l_{1i} = \prod_{1 \leq j \leq n, j \neq i} \frac{(q_1 - s_1^j)}{(s_1^i - s_1^j)} \quad \forall 1 \leq i \leq n \quad (4)$$

$$l_{2i} = \prod_{1 \leq j \leq n, j \neq i} \frac{(q_2 - s_2^j)}{(s_2^i - s_2^j)} \quad \forall 1 \leq i \leq n \quad (5)$$

$$l_{3i} = \prod_{1 \leq j \leq n, j \neq i} \frac{(q_3 - s_3^j)}{(s_3^i - s_3^j)} \quad \forall 1 \leq i \leq n \quad (6)$$

The effective multiplier at each point is calculated as average of these three multipliers.

$$\lambda_i = \frac{l_{1i} + l_{2i} + l_{3i}}{3} \quad \forall 1 \leq i \leq n \quad (7)$$

The location of the user is estimated as:

$$(X, Y) = \left(\sum_{i=1}^n x_i \lambda_i, \sum_{i=1}^n y_i \lambda_i \right) \quad (8)$$

The likelihood, L of the user's location being (X, Y) is given by:

$$L(q_1, q_2, q_3) = \sum_{i=1}^n P_i(q_1, q_2, q_3) \lambda_i(s) \quad (9)$$

where $P_i(q_1, q_2, q_3)$ is the probability that the user is located at the i^{th} point of the given cluster. Since the base stations operate on non-overlapping channels, their signal strength probability distribution functions can be assumed to be independent of each other. Based on this assumption $P_i(q_1, q_2, q_3)$ can be expressed as:

$$P_i(q_1, q_2, q_3) = P_i^1(s = q_1) \cdot P_i^2(s = q_2) \cdot P_i^3(s = q_3) \quad (10)$$

where $P_i^j(s = q_j)$ can be calculated from the frequency distribution table of the i^{th} point for j^{th} base station as described in the section 3.1.

3.2.4 Choosing the location with the highest likelihood

From the previous subsection, m probable candidate locations $\{(X_i, Y_i)\}_{i=1}^m$ and the corresponding likelihoods $\{L_i\}_{i=1}^m$ for each of the m clusters are obtained. The candidate with the highest likelihood is chosen to be the estimate for the user's current location.

4. EXPERIMENTAL EVALUATION

In this section, first the experimental setup, which includes a description of the testbed and the various configurable parameters of the LOCATOR system, is described. Next, the effect of the various parameters on the performance of LOCATOR is discussed. Finally, a performance evaluation of LOCATOR against two other RF-based location estimation systems namely RADAR [9] and Horus [12] is described. For a fair performance evaluation, all three systems were tested under the same environmental conditions.

4.1 Experimental Setup

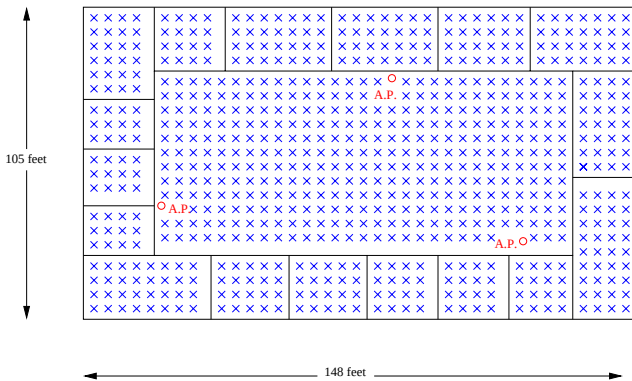


Figure 3: Experimental testbed

The testbed (figure 3), a school corridor surrounded by classrooms, is used for the performance evaluation of LOCATOR. It covers an area of approximately 15,000 square

feet and has three 802.11b wireless access points of Cisco 1200 series deployed at the positions shown in the figure. The access points operate on channels 3, 6 and 9, which are non-overlapping and therefore, do not cause mutual interference. The crosses shown in the above figure indicate the locations where signal strength samples were collected while building the radio map database.

The separation between the crosses defines the granularity of the radio map and is denoted by g . Another parameter of interest is the number of candidate points (k) to be considered for clustering in step 2 of the location estimation phase. This number defines the value for k in the k -best search that is performed in step 1. Finally, there is a parameter γ , which is the ratio of the maximum diameter of a cluster and the granularity of the radio map. Thus, the maximum diameter of the cluster that is used in step 2 of the location estimation phase is given by $\gamma \cdot g$.

For building the radio map database, a large number of signal strength samples are recorded at each location. The number of signal strength samples recorded should be large enough so that the signal strength probability distribution function for each location can be accurately captured. For our experimental testbed, taking 150 samples at each location proved to be sufficient. Using these samples, the frequency distribution table is constructed for each of the locations. Also, the mean signal strength values at each of the locations is also recorded. The mean signal strength values are used in step 1 of location estimation phase for k -best search while the frequency distribution tables are used in step 2 for selecting the most probable location estimate for each cluster.

4.2 Effect of Parameters on Performance

LOCATOR has three control parameters: (a) Granularity of radio map denoted by g , (b) Number of candidates selected for clustering denoted by k and (c) Ratio of the maximum diameter of a cluster and the granularity of the radio map denoted by γ . In this section, the effect of variation of each of these control parameters on the performance of LOCATOR is studied. For quantifying the performance of LOCATOR, the error distance at 90 percentile is used as the performance metric i.e. for instance, if 100 observations for location estimation are made and 90 of these have error distance, which is the difference between actual position and estimated position, with in 3 feet, then error distance at 90 percentile is 3 feet.

First the effect of varying the number of candidates (k) used for clustering, on the performance of LOCATOR is studied. While varying k , g is kept fixed at 10 feet and γ at 3. As observed from figure 4, as k increases the 90 percentile error distance initially decreases and then becomes almost constant. This can be explained as follows. Initially when k is small, certain points that are physically closer to the actual user location but not that close in the signal space are not selected in the k -best search. As a result, the performance is poor. As the value of k is increased, eventually all points that are physically close to that user's current location are selected which naturally improves the accuracy. But as k is increased further, there is no improvement in the performance because all the neighboring points of the user's current location get selected after a certain threshold value of k and thus, cause saturation in accuracy.

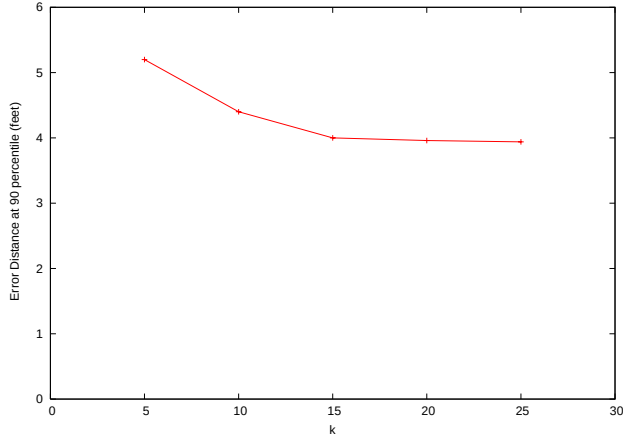


Figure 4: Effect of number of candidates, k , used for clustering

From figure 4, 15 seems to be the optimum value for k and would be used while comparing LOCATOR's performance against other radio frequency based systems in the next section.

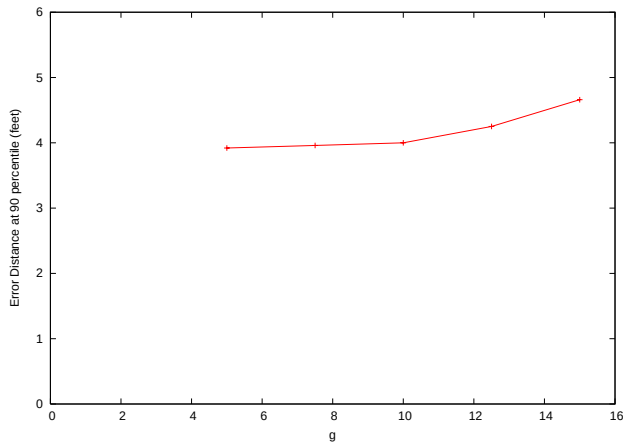


Figure 5: Effect of granularity, g

In figure 5, the effect of the granularity (g) on the performance is studied while keeping k at 15 and γ at 3. As expected, the 90 percentile error distance increases as the granularity of the radio map decreases i.e. g increases. From the figure, it is observed that initially the error distance increases at a slower rate than the rate at which g is increased. The initial slow rate of increase of the error distance can be attributed to the use of interpolation while choosing most probable locations in step 3 of location estimation phase. If interpolation were not to be used, the error distance increases linearly with the granularity. But as the granularity decreases, the cost in terms of time required to build the radio map database i.e. the time required for the radio map building phase increases. We would use the value of granularity as 10 feet as it gives the best mix of performance in terms of accuracy and the time required for radio map building phase.

Finally, the effect of variation of the ratio of the maximum diameter of a cluster and the granularity of the radio map

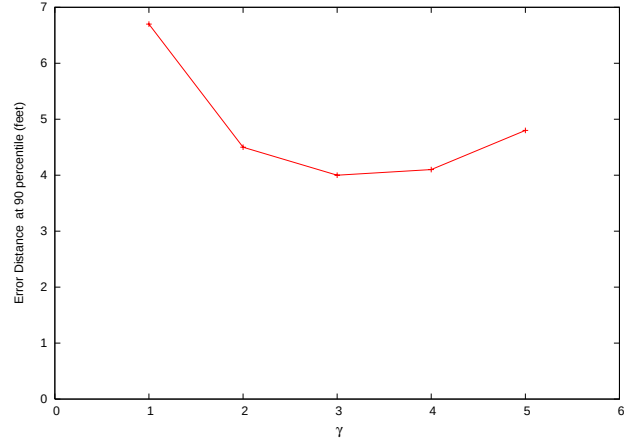


Figure 6: Effect of ratio of the maximum diameter of a cluster and the radio map granularity

(γ) on the performance of LOCATOR is studied. In figure 6, the value of γ is varied from 1 to 5 while keeping k at 15 and g at 10 feet. As observed from the figure, initially when γ is 1, the error is large which is due to the fact that there will be no clustering of points for γ equal to 1. As γ is increased, there is an increase in the performance due to clustering and interpolation of the points in the cluster. But as γ is increased further, a drop in the performance is observed. This drop in the performance can be attributed to the fact that a large value of γ can result in formation of very large clusters that comprise of points that are spaced out widely. Since all the points in a cluster are used for interpolation to estimate the most probable candidate for each cluster, as discussed in section 3.3.2.3, therefore, these distant points lead to greater errors in location estimation.

4.3 Performance evaluation against other RF-based systems

In this section, the performance of LOCATOR is compared against other radio frequency (RF) based location estimation systems, namely, RADAR[9] and Horus[12].

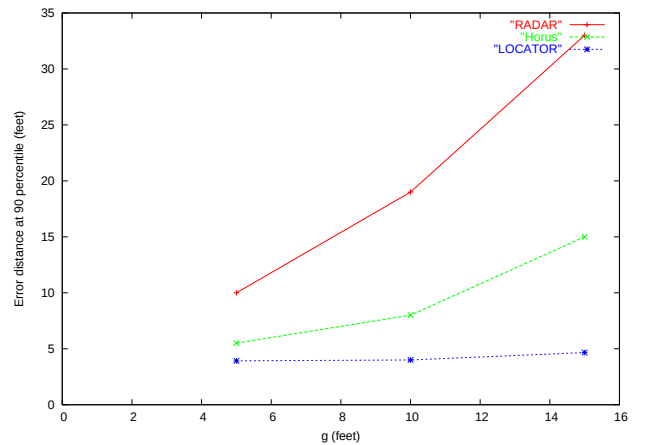


Figure 7: Performance comparison of RF-based systems on the basis of radio map granularity

First, a performance comparison of the three systems is studied on the basis of granularity of the radio map. During the evaluation, k has been kept fixed at 15 and γ at 3 for the LOCATOR system. As observed in figure 7, LOCATOR outperforms both RADAR and Horus by a significant margin and as the granularity of the radio map decreases i.e g increases, the difference in the performance becomes even more prominent. This can be attributed to the interpolation technique employed by LOCATOR to select the most probable user location for each cluster and then select the candidate with the highest likelihood as the estimate for the user location. Horus performs better than RADAR since it uses its Joint Clustering technique[12] rather than simply doing a k -nearest neighbor search as in the case of RADAR. Since neither Horus nor RADAR use interpolation, therefore the accuracy of these systems is severely limited by the granularity of the radio map. Unlike RADAR and Horus, its again due to interpolation that the performance of LOCATOR degrades gracefully as the granularity is increased.

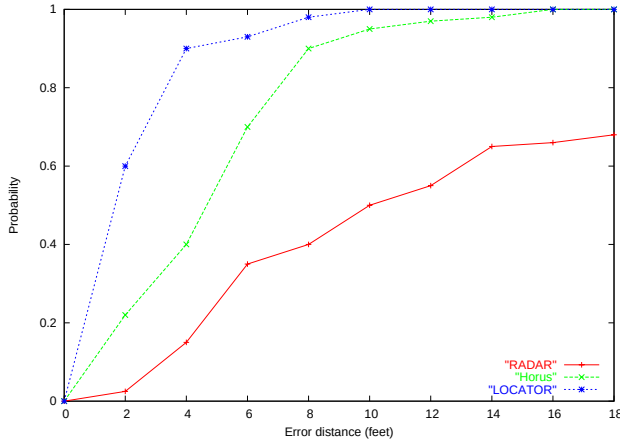


Figure 8: Error distance CDF for various RF-based systems

Next, the cumulative distribution functions (CDFs) of the error distance are presented for each of the three RF-based systems in figure 8. As observed from the figure, LOCATOR gives 90% accuracy to within 4 feet while Horus and RADAR give only 40% and 15% accuracy to within 4 feet respectively for the testbed setup that we have used.

Finally, the performance of the three systems is analyzed on the basis of the computation cost for processing location queries. For a fair analysis, all the three systems were tested on the same machine and under similar load conditions. For the purpose of comparison, the *system time* output of the Unix *time* command was used to measure the response time for location queries. Figure 9 has been plotted by taking an average of 500 location estimation queries for each of the three systems. As seen from the figure, LOCATOR significantly outperforms Horus. This can be attributed to LOCATOR's location-based clustering of the radio map due to which the entire radio map database is not searched for answering location queries, thus, leading to lower computation overhead. The performance of RADAR is just marginally worse than that of LOCATOR. This phenomenon can be attributed to two mutually nullifying factors in RADAR's search procedure. The RADAR system searches the entire radio map database for answering location queries and thus,

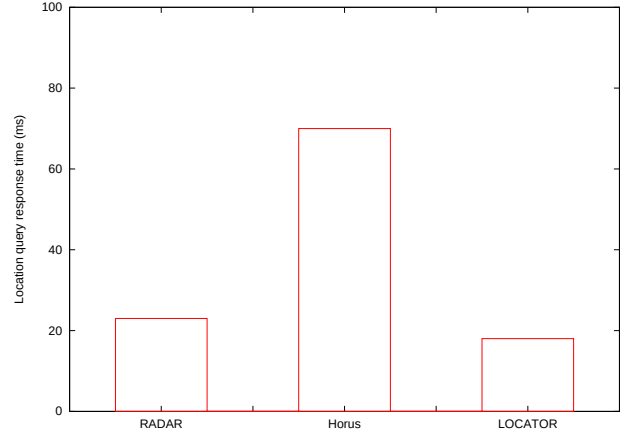


Figure 9: Computation Overhead

incurs a huge cost. This high computation cost of searching the entire radio map database is balanced out by RADAR's use of a simple, low computation cost k -nearest neighbor search for answering location queries. In conclusion, as the size of the radio map increases, the performance of LOCATOR will improve further due to its location-based clustering of the radio map database.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented LOCATOR, a new radio frequency based location estimation system, and its performance evaluation against other RF-based systems. The effect of various control parameters on LOCATOR's performance has also been analyzed.

LOCATOR incorporates several novel ideas that lead to its superior performance. In the radio map building phase, the radio map is fragmented into location-based clusters that reduce the computation cost in the location estimation phase by a significant factor and also makes LOCATOR scalable and thus, deployable for larger wireless networks. The use of signal strength probability distribution functions, clustering and interpolation in the location estimation phase lead to a more accurate user location estimation.

The implementation results of LOCATOR on our 802.11b experimental testbed show an accuracy of 90% for error distances within 4 feet. Also, LOCATOR is a generic system that can be used for other wireless network technologies like Bluetooth etc.

In this paper, we have considered only one kind of interpolation scheme namely Lagrange interpolation. In our future work, we would like to study the effect of using other interpolation schemes and analyze the trade-off between the increase in computation cost and the increase in accuracy due to different interpolation schemes. Since the requirement, in terms of accuracy, for users' location information varies from application to application, we would also like to explore the possibility of making LOCATOR adaptive to the context in which it is being deployed. For instance, if some location-based service requires highly precise user location information then LOCATOR would switch to an expensive but more accurate interpolation scheme to provide more accurate user location.

6. REFERENCES

- [1] P. Enge and P. Misra. *Special issue on GPS: The Global Positioning System*. Proceedings of the IEEE, pages 3-172, January 1999.
- [2] S. Tekinay. *Special issue on Wireless Geolocation Systems and Services*. IEEE Communications Magazine, April 1998.
- [3] R. Want, A. Hopper, V. Falco, and J. Gibbons. *The Active Badge Location System*. ACM Transactions on Information Systems, 10(1):91-102, January 1992.
- [4] R. Azuma. *Tracking requirements for Augmented Reality*. Communications of the ACM. Vol. 36, No. 7, pp: 50-51, July 1993
- [5] W. Stallings. *Wireless Communications and Networks*. Prentice Hall, first edition, 2002.
- [6] The Institute of Electrical and Electronics Engineers, Inc. IEEE Standard 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. 1999.
- [7] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. *The Cricket Location-Support system*. In 6th ACM MOBICOM, Boston, MA, August 2000.
- [8] R. J. Orr and G. D. Abowd. *The Smart Floor: A Mechanism for Natural User Identification and Tracking*. In Conference on Human Factors in Computing Systems (CHI 2000), pages 1-6, The Hague, Netherlands, April 2000.
- [9] P. Bahl and V. N. Padmanabhan. *RADAR: An In-Building RF-based User Location and Tracking System*. In IEEE Infocom 2000, volume 2, pages 775-784, March 2000.
- [10] P. Castro, P. Chiu, T. Kremenek, and R. Muntz. *A Probabilistic Location Service for Wireless Network Environments*. Ubiquitous Computing 2001, September 2001.
- [11] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen. *A Probabilistic Approach to WLAN User Location Estimation*. International Journal of Wireless Information Networks, 9(3), July 2002.
- [12] M. Youssef, A. Agrawala, and A. U. Shankar. *WLAN Location Determination via Clustering and Probability Distributions*. In IEEE PerCom 2003, March 2003.