

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/228964423>

# Location-clustering techniques for WLAN location determination systems

Article · January 2006

DOI: 10.2316/Journal.202.2006.3.202-1772

---

CITATIONS

20

---

READS

994

2 authors:



[Moustafa Youssef](#)

Egypt-Japan University of Science and Technol...

219 PUBLICATIONS 5,341 CITATIONS

[SEE PROFILE](#)



[Ashok Agrawala](#)

University of Maryland, College Park

298 PUBLICATIONS 5,827 CITATIONS

[SEE PROFILE](#)

All content following this page was uploaded by [Ashok Agrawala](#) on 10 January 2017.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

# LOCATION-CLUSTERING TECHNIQUES FOR WLAN LOCATION DETERMINATION SYSTEMS

Moustafa Youssef\*, Ashok Agrawala

Department of Computer Science

University of Maryland at College Park

College Park, MD 20742

{moustafa, agrawala}@cs.umd.edu

## Abstract

We present location clustering as a technique to significantly reduce the computational requirements of WLAN location determination systems. We provide two algorithms, namely, Joint Clustering and Incremental Triangulation, and describe their tradeoffs between computational cost and location determination accuracy. Both techniques reduce computational cost by more than an order of magnitude, allowing non-centralized implementation on mobile clients and enabling new context-aware applications. We present a performance comparison of the two techniques in an actual testbed implementation.

---

\* Also affiliated with Alexandria University, Egypt.

## Keywords

Energy efficient WLAN location determination, location-aware systems, radio map locations clustering, user positioning, wireless LAN.

## 1. Introduction

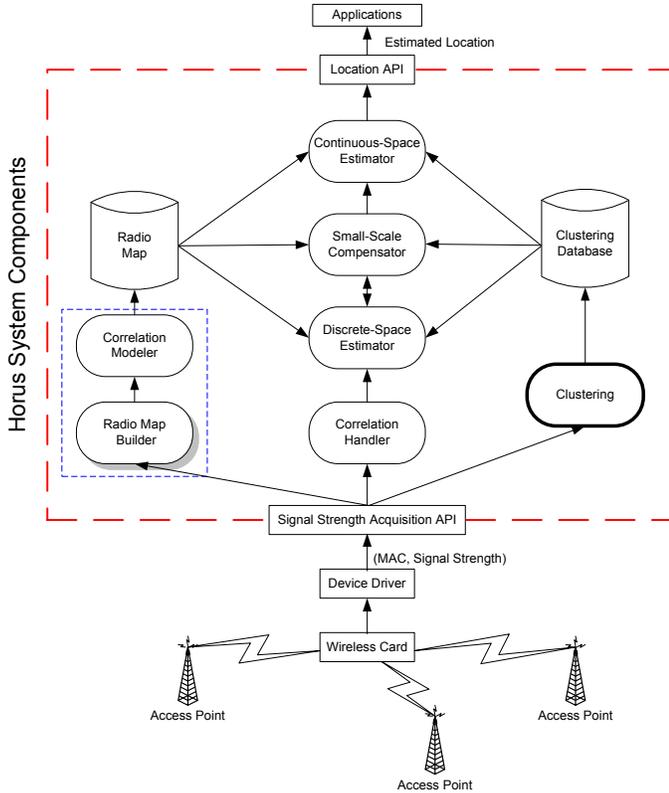
As ubiquitous computing becomes more popular, the importance of context-aware applications increases [1]. This in turn fuels the need to determine the user location, with which the system can provide location-specific information and services.

WLAN location determination systems [2–11] use an underlying wireless data network, such as 802.11, to estimate the user location. WLAN-based techniques provide more ubiquitous coverage and do not require additional hardware for user location determination, thereby enhancing the value of the wireless data network.

WLAN location determination systems usually work in two phases: *offline* training phase and *online* location determination phase. During the offline phase, the signal strength received from the access points, at selected locations in the area of interest, is tabulated resulting in a so-called *radio map*. During the location determination phase, the signal strength samples received from the access points are used to “search” the radio map to estimate the user location.

Since mobile devices are energy-constrained, it is important to reduce the computational requirement of the location determination system to extend the life of their batteries.

In this paper, we introduce clustering of radio map locations as an approach to reduce the computational requirements of the location determination algorithms. The proposed algorithms reduce computational requirements by running the location determination algorithm for only a small subset (a cluster)



**Figure 1. *Horus* Components:** the arrows show information flow in the system. Shaded block represent modules used during the offline phase. In this paper, we describe the clustering component of the *Horus* system (shown in thick lines).

of the radio map locations, saving significant amount of computation. Achieving such clustering is challenging given the noisy characteristics of the wireless channel. We describe two general approaches to clustering: explicit clustering and implicit clustering, both applicable to any of the current radio map based approaches. We also describe the tradeoffs between the computational cost and location determination accuracy for the proposed techniques.

Results from 802.11-equipped iPAQ implementations in the context of the Horus probabilistic WLAN location determination system [2–5] show that both algorithms reduce computational cost by more than an order of magnitude without sacrificing accuracy.

The rest of the paper is organized as follows. In the next section, we present a brief introduction

of the *Horus* WLAN location determination system. Section 3 presents the noisy characteristics of the wireless channel that make the clustering problem challenging. In Section 4 we present the details of radio map construction and location estimation with the explicit clustering and implicit clustering techniques. In Section 5, we describe the evaluation of the techniques in the context of the *Horus* WLAN location determination system and the obtained results. Finally, we discuss related work in Section 6 and conclude the paper in Section 7.

## 2. The Horus System

*Horus* [2–5] is a probabilistic location determination system. The main goal of the system is to identify the noisy characteristic of the wireless channel and to develop techniques to handle them. Fig. 1 shows the components of the *Horus* system. The system uses the signal strength information returned from different access points to infer the user location and to provide an API for the user applications to use the system functionality.

The system works in two phases:

1. *Offline phase*: to build the radio map, cluster radio map locations, and do other preprocessing of the signal strength models.
2. *Online Phase*: to estimate the user location based on the received signal strength from each access point and the radio map prepared in the offline phase.

The radio map stores the distribution of signal strength received from each access point at each location.

The *Clustering* module is used to group radio map locations based on the access points covering them.

---

**Alg. 1**  $x = \text{Horus\_GetLocation}(s, \mathbb{X}, P\_RM)$ 

---

**Require:** $s$  : Measured signal strength vector from  $k$  access points ( $s = (s_1, \dots, s_k)$ ). $\mathbb{X}$  : Radio map locations. $RM$  : A radio map based function, where  $P\_RM(s_a, a, x)$  returns the probability of receiving signal strength  $s_a$  from access point  $a$  at location  $x \in \mathbb{X}$ .**Ensure:**The location  $x \in \mathbb{X}$  that maximizes  $P(x/s)$ .1:  $\text{Max} \leftarrow 0$ 2: **for**  $l \in \mathbb{X}$  **do**3:  $P \leftarrow \prod_{i=1}^k P\_RM(s_i, i, l)$ 4: **if**  $P > \text{Max}$  **then**5:  $x \leftarrow l$ 6:  $\text{Max} \leftarrow P$ 7: **end if**8: **end for**

---

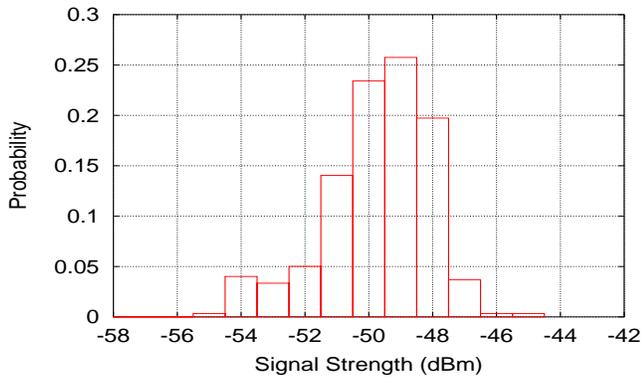
Clustering is used to reduce the computational requirements of the system and, hence, conserve power.

This module is the subject of this paper.

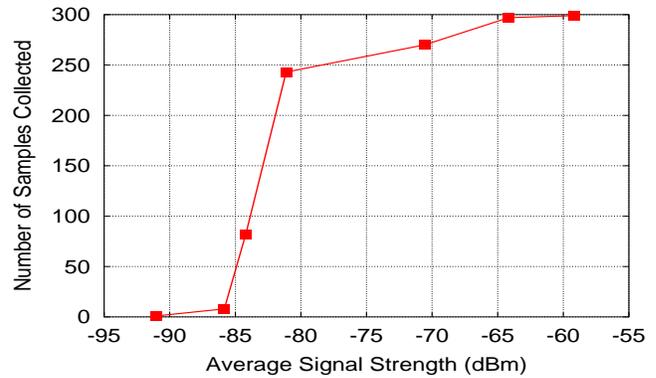
The *Discrete Space Estimator* module returns the radio map location that has the maximum probability given the received signal strength vector from different access points. An outline of the algorithm used is given in Algorithm 1. We use this algorithm as the baseline for comparing the performance of the proposed clustering techniques.

The *Small-Scale Compensator* module handles the small-scale variation characteristics of the wireless channel [4].

The *Continuous Space Estimator* takes as an input the discrete estimated user location, one of the radio map locations, and returns a more accurate estimate the user location in the continuous space [5].



(a) Histogram of the signal strength of an access point



(b) Relation between the average signal strength of an access point and the percentage of samples received from it.

**Figure 2. Wireless channel characteristics.**

### 3. Wireless Channel Characteristics

Fig. 2(a) gives a typical example of the normalized histogram of the signal strength received from an access point at a fixed location. People moving in the environment, doors opening and closing, and other changes in the environment can explain such temporal changes shown in the figure.

We also performed an experiment to test the behavior of access points with different average signal strength at the same location. During this experiment, we sampled the signal strength from each access point at the rate of one sample per second. Fig. 2(b) shows the relation between the average signal strength received from an access point and the percentage of samples we receive from it during a period of 5 minutes. The figure shows that the number of samples collected from an access point is a monotonically increasing function of the average signal strength. Assuming a constant noise level, the higher the signal strength, the higher the signal to noise ratio and the more probable it becomes that the 802.11b card will identify the existence of a frame.

To summarize, figure 2 highlights the characteristics of the wireless channel: (1) At a fixed location, the signal strength received from an access point varies with time. (2) The number of access points

covering a location varies with time.

## 4. Radio Map Locations Clustering

### 4.1. Approach

We define a *cluster* as a set of locations sharing a common set of access points. We call this common set of access points the *cluster key*. The problem can be stated as: Given a location  $x$ , we want to determine the cluster to which  $x$  belongs. The noisy characteristics of the wireless channel described in Section 3 make clustering a challenging problem because the number of access points covering a location varies with time.

We present two approaches:

- *Explicit clustering*: Here the system must determine the clusters during the offline training phase as a separate step.
- *Implicit clustering*: Here, no special processing is performed in the offline phase. However, during the location determination phase, the system performs clustering implicitly.

The details of the algorithms are given next.

### 4.2. Explicit Clustering

One way to do clustering is to group locations according to the access points that cover them. i.e. two locations  $x_1$  and  $x_2$  are placed in the same cluster *iff* the set of access points covering these locations are identical. However, this approach for clustering has problems when applied in a real environment. As shown in Fig. 2(b), an access point may be missing from some of the samples and, therefore, using the

entire set of access points that cover a location for clustering may fail to find the correct cluster due to the missing access point.

Instead, we use a subset of this set containing only  $q$  elements and the problem becomes: Given a number  $q$ , we want to put all the locations that share  $q$  access points in one cluster. Therefore, we have 2 sub-problems: (1) How to determine the value of  $q$ ? and (2) Which  $q$  access points to choose for clustering?

For the first sub-problem, we need to choose  $q$  such that all locations are covered by at least  $q$  access points most of the time. This factor lessens the effect of variability in the number of access points with time. This suggests that the value of  $q$  should be less than or equal to the minimum number of access points covering any location in the radio map. Moreover, we need a value for  $q$  that distributes locations evenly between the clusters to reduce the required computations. Experiments showing the effect of the parameter  $q$  on performance are given in Section 5.

The solution for sub-problem 2 is related to the solution of sub-problem 1. If the number of access points covering a location is varying with time, which access points should we choose? Intuitively, we should choose the access points that appear most of the time in the samples. Fig. 2(b) suggests that we should choose to use the  $q$  access points with the largest signal strength values at each location.

To summarize, for a given location  $x$ , we use the set of the  $q$  strongest access points covering this location to determine the cluster to which it belongs. Therefore, the cluster key in the *Explicit Clustering* approach is the set of the  $q$  access points used to group the locations in this cluster.

We call the modified location determination algorithm, that uses the *Explicit Clustering* technique, the *Joint Clustering* algorithm (Algorithm 2). The algorithm is identical to the previously described *Horus* algorithm with the exception of reducing the radio-map space to a single cluster.

---

**Alg. 2**  $x=JC\_GetLocation(s, \mathbb{X}, RM, Cluster, q)$ 

---

**Require:**

- $s$  : Measured signal strength vector ( $s = (s_1, \dots, s_k)$ ).
- $RM$  : Radio map.
- $Cluster$  : Clustering function where  $Cluster(k)$  is the set of radio-map locations whose key is  $k$ .
- $q$  : Number of access points to use in clustering.

**Ensure:**Estimated location  $x$ .

- 1:  $OrderedS \leftarrow s$  sorted in a descending order.
  - 2:  $CandidateList \leftarrow Cluster(OrderedS(1..q))$
  - 3:  $x \leftarrow GetLocation(s, CandidateList, RM)$ .  
{Get the candidate location using standard algorithms}
- 

### 4.3. Implicit Clustering

One can look at the clustering problem from a different viewpoint. Each access point defines a subset of the radio map locations that are covered by this access point. These locations can be viewed as a cluster of locations whose key is the access point covering the locations in this cluster. If during the location determination phase we use the access points incrementally, one after the other, then starting with the first access point, we restrict our search space to the locations covered by this access point. The second access point chooses only the locations in the range of the first access point and covered by the second access point and so on, leading to a multi-level clustering process.

Notice that no preprocessing is required in the offline training phase. During the online phase, a location  $x$  belongs to a cluster whose key is access point  $a$  if there is information about access point  $a$  at location  $x$  in the radio map.

We call the modified location determination algorithm, that uses the *Implicit Clustering* technique, the *Incremental Triangulation* algorithm (Algorithm 3). The algorithm works as follows. Given a sequence of observations from each access point, we start by sorting the access points in descending order according to the average received signal strength. For the first access point, the one with the strongest

average signal strength, we calculate the probability of each location in the radio map set given the observation sequence from this access point alone. This gives us a set of candidate locations (locations 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 our location estimate, after consulting only one access point. If this is not the case, we go to the next access point in the sorted access point list. For this access point, we repeat the same process again, but only for the set of candidate locations obtained from the first access point.

---

**Alg. 3**  $x = \text{IT\_GetLocation}(s, \mathbb{X}, RM, \text{Threshold})$

---

**Require:**

- $s$  : Measured signal strength vector ( $s = (s_1, \dots, s_k)$ ).
- $\mathbb{X}$  : Radio map locations.
- $RM$  : Radio map.
- $\text{Threshold}$  : Stopping threshold.

**Ensure:**

Estimated location  $x$ .

- 1:  $\text{OrderedS} \leftarrow s$  sorted in a descending order.
  - 2:  $\text{CandidateList} \leftarrow \mathbb{X}$
  - 3:  $\text{CurrentAP} \leftarrow 1$ .
  - 4:  $\text{Done} \leftarrow \text{false}$ .
  - 5: **while** Not  $\text{Done}$  **do**
  - 6:   **for**  $l \in \text{CandidateList}$  **do**
  - 7:      $\bar{s} \leftarrow \text{OrderedS}(1..\text{CurrentAP})$ .
  - 8:      $p(l) \leftarrow$  probability of location  $l$  using  $\text{GetLocation}(\bar{s}, \text{CandidateList}, RM)$ .  
      {Get the probability using standard algorithms}
  - 9:   **end for**
  - 10:   sort  $p$  in a descending order and sort  $\text{CandidateList}$  accordingly.
  - 11:   **if**  $\left(\frac{p(\text{CandidateList}(1)) - p(\text{CandidateList}(2))}{p(\text{CandidateList}(1))} > \text{Threshold} \parallel \text{CurrentAP} == k\right)$  **then**
  - 12:      $\text{Done} = \text{true}$
  - 13:   **else**
  - 14:      $\text{CurrentAP} = \text{CurrentAP} + 1$ .
  - 15:      $\text{CandidateList} \leftarrow$  Elements with non-zero probability of  $\text{CandidateList}$ .
  - 16:   **end if**
  - 17: **end while**
  - 18:  $x \leftarrow \text{CandidateList}(1)$
- 

This process of calculating the probabilities and determining the significance of the most probable

location is repeated incrementally, for each access point in order, until the location can be estimated or all access points are consulted. In the latter case, the algorithm returns the most probable location in the candidate list that remains after consulting all the access points. We call our approach the *Incremental Triangulation* technique.

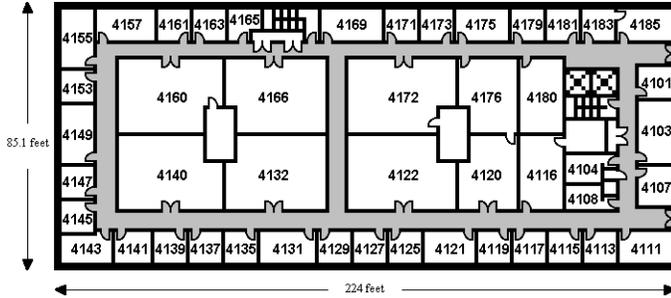
#### 4.4. Discussion

Both clustering techniques reduce the search space and thus lead to a reduction of the computational cost of location determination techniques employing these clustering approaches. Moreover, using clustering helps in scaling the system to a larger coverage area.

The *Explicit Clustering* divides the radio-map space  $\mathbb{X}$  in a number of flat clusters. However, the *Implicit Clustering* tries to use the access points incrementally, one after the other, until it can estimate the location with certain accuracy leading to clustering at multi-levels and hence to a more reduced search space than the *Explicit Clustering* approach, and hence fewer number of operations, on the average, per sample. However, treating each access point incrementally, instead of using the joint distribution, leads to the loss of some information and thus the accuracy of the *Implicit Clustering* is lower than the *Explicit Clustering* technique.

## 5. Experimental Evaluation

In this section, we discuss the experimental testbed and evaluate the performance of the Joint Clustering and Incremental Triangulation techniques.



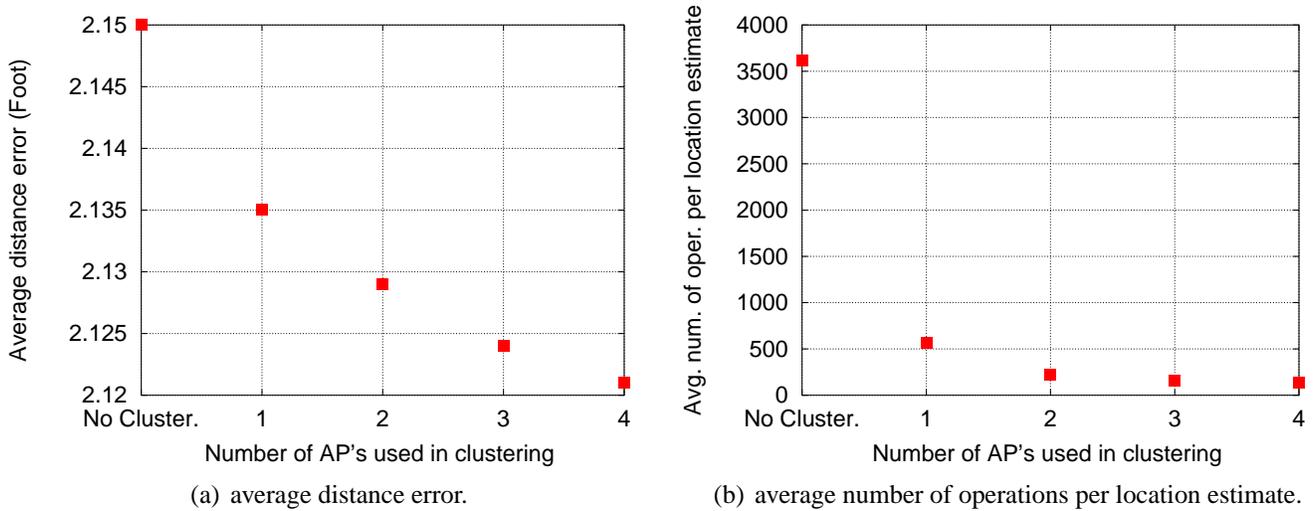
**Figure 3. Floor plan for the testbed. Readings were collected in the corridors and inside the rooms.**

### 5.1. Experimental Testbed

We collected 200 samples at each radio map location, one sample every 100 milliseconds. The cards used were Lucent Orinoco silver NICs supporting up to 11 Mbit/s data rate [12]. *To test the performance of the system, we used an independent test set that was collected on different days, time of day, and by different persons than the training set.*

We performed our experiment in the south wing of the fourth floor of the A. V. William’s building in the University of Maryland at College Park. The layout of the floor is shown in Fig. 3. The wing has a dimension of 224 feet by 85.1 feet. The technique was tested in the University of Maryland wireless network using *Cisco* access points. The entire wing is covered by 21 access points. The radio map has 110 locations along the corridors and 62 locations inside the rooms. On the average, each location is covered by 6 access points. The *Horus* system was running on *Windows XP* professional operating system.

To measure the performance of the proposed techniques, we used two metrics: (1) *Accuracy*: This measure is defined as the percentage of time in which the technique gives the correct location estimate within a certain distance and (2) the *Number of operations per location estimate*: This measure is defined as the total number of operations (multiplications) performed for a single location estimate. This is



**Figure 4. Effect of the parameter  $q$  on the performance of the  $JC$  technique.**

important in minimizing computation time, but more so in minimizing the power consumption.

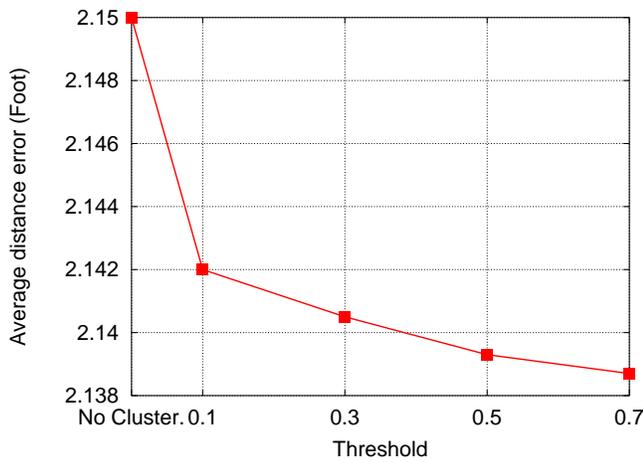
## 5.2. Joint Clustering Technique

Fig. 5.2 shows the effect of the parameter  $q$  on the average distance error. We can see that as  $q$  increases, the accuracy of the system slightly increases. Since, clustering reduces the search space, clustering can enhance the system accuracy.

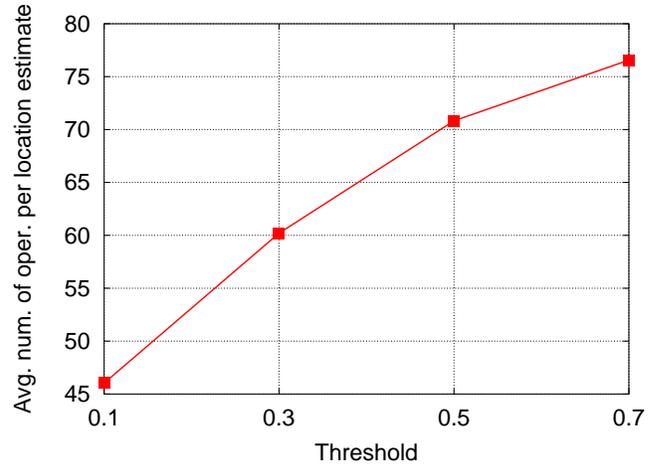
Fig. 4(b) shows the effect of the parameter  $q$  on the average number of operations per location estimate. As  $q$  increases, the average number of operations per location estimate decreases. For this range of  $q$  values, as  $q$  increases, the average cluster size decreases and hence the system uses a fewer number of operations per location estimates.

## 5.3. Incremental Triangulation Technique

Figures 5(a) and 5(b) show the effect of the parameter  $Threshold$  on the performance. For small values of the  $Threshold$  parameter, the decision is taken quickly after examining a small number of access



(a) average distance error.



(b) average number of operations per location estimate.

**Figure 5. Effect of the parameter *Threshold* on the performance of the *IT* technique.**

points. As the threshold value increases, more access points are consulted to reach a decision. As the number of access points consulted increases, the number of operations per location estimate increases and the so does the accuracy.

## 6. Related work

Radio map-based techniques can be categorized into two broad categories: deterministic techniques [6–8] and distribution-based techniques [2–5, 9–11]. Our work lies in the second category. However, none of the previous systems take into account the computational burden of the location determination algorithm. Our work is unique in introducing clustering of radio map locations as an approach to reduce the computational requirements of the location determination techniques, enhance accuracy, and increase the scalability of the system. Therefore, our location determination system combines both merits: high accuracy with significant reduction in computational cost.

## 7. Conclusions

In this paper, we presented clustering of radio map locations as an approach to reduce the computational requirements of the location determination algorithms and achieve scalability. We also showed that clustering of radio map locations is a challenging problem with the noisy characteristic of the wireless channel. We describe two general approaches to clustering: *Explicit Clustering* and *Implicit Clustering*. The results show that using clustering reduces the average number of operations per location estimate by more than an order of magnitude. The *Explicit Clustering* technique gives slightly better accuracy than the *Implicit Clustering* technique. However, The average number of operations performed per location estimate for the *Implicit clustering* technique is much lower than the corresponding number of the *Explicit Clustering* technique.

Such energy saving allows the system to be implemented on energy-constrained mobile devices and thus increases the scalability of the system in terms of the number of supported users. The proposed clustering techniques can be applied to all the current WLAN location determination systems to reduce their computational cost and enhance their accuracy.

## References

- [1] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Tech. Rep. Dartmouth Computer Science Technical Report TR2000-381, 2000.
- [2] M. Youssef, A. Agrawala, and A. U. Shankar, "WLAN Location Determination via Clustering and Probability Distributions," in *IEEE PerCom 2003*, March 2003.
- [3] Moustafa Youssef and Ashok Agrawala, "Handling samples correlation in the horus system," in *IEEE Infocom*, March 2004.
- [4] M. Youssef and A. Agrawala, "Small-Scale Compensation for WLAN Location Determination Systems," in *IEEE WCNC 2003*, March 2003.
- [5] Moustafa Youssef and Ashok Agrawala, "Continuous space estimation for wlan location determination systems," in *IEEE International Conference on Computer Communications and Networks*, October 2004.
- [6] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," in *IEEE Infocom 2000*, March 2000, vol. 2, pp. 775–784.

- [7] A. Smailagic, D. P. Siewiorek, J. Anhalt, D. Kogan, and Y. Wang, "Location Sensing and Privacy in a Context Aware Computing Environment," *Pervasive Computing*, 2001.
- [8] P. Krishnan, A.S. Krishnakumar, Wen-Hua Ju, Colin Mallows, and Sachin Ganu, "A system for lease: Location estimation assisted by stationary emitters for indoor rf wireless networks," in *IEEE Infocom*, March 2004.
- [9] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Location Service for Wireless Network Environments," *Ubiquitous Computing 2001*, September 2001.
- [10] 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*, vol. 9, no. 3, July 2002.
- [11] A. M. Ladd, K. Bekris, A. Rudys, G. Marceau, L. E. Kavraki, and D. S. Wallach, "Robotics-Based Location Sensing using Wireless Ethernet," in *8th ACM MOBICOM*, Atlanta, GA, September 2002.
- [12] "<http://www.orinocowireless.com>," .



*Moustafa Youssef* is a faculty research associate in the Department of Computer Science at the University of Maryland at College Park. He received his B.Sc. and M.Sc. in Computer Science from Alexandria University, Egypt in 1997 and 1999 respectively and the Ph.D. degree in computer science from University of Maryland in 2004. His research interests include location determination technologies, pervasive computing, energy-aware computing, sensor networks, and protocol modeling. Dr. Moustafa is a life fellow for the Egyptian Society for Talented, an elected member of the honor society Phi Kappa Phi, among others. He is a member of various professional societies such as IEEE, IEEE Computer Society, IEEE Communication Society and ACM Sigmobile. Dr. Moustafa is the recipient of the 2003 University of Maryland Invention of the Year award for his Horus work.



*Agrawala Agrawala* is a professor at the University of Maryland at College Park. In 2001, he started the Maryland Information and Network Dynamics (MIND) Lab which carries out research and development activities in partnership with the industry. He received a BE degree in 1963 and a ME in 1965 from the I.I.Sc, Bangalore; and a Master of Arts and a Ph.D. degree in Applied Mathematics from Harvard University in 1970. Prof. Agrawala is the author of seven books, 6 patents (awarded or pending), and over 240 papers and is a recognized authority in the research and use of the management of time in real-time processing and clock synchronization applications. He has developed a few location determination techniques and several other innovative technologies for systems and networks which are in different stages of deployment. Prof. Agrawala is a Fellow of the IEEE and Senior Member of the ACM.