A Statistical Indoor Localization Method for Supporting Location-based Access Control (Invited Paper)

Chunwang Gao, Zhen Yu, Yawen Wei, Steve Russell, and Yong Guan
Department of Electrical and Computer Engineering
Iowa State University, Ames, IA 50011, USA
gaojerry@gmail.com and {yuzhen, weiyawen, sfr, yguan}@iastate.edu

ABSTRACT

Location awareness is critical for supporting location-based access control (LBAC). The challenge is how to determine locations accurately and efficiently in indoor environments. Existing solutions based on WLAN signal strength either cannot provide high accuracy, or are too complicated in general indoor environments. In this paper, we propose a statistical indoor localization method for supporting location-based access control. In an offline phase, we fit a LOESS [3, 4, 16] local regression model on a training set to build a radio map containing the distribution of signal strength. In an online phase, we estimate locations using Maximum Likelihood Estimation (MLE) [7, 8, 9] based on the measured signal strength and the stored distribution. A Bootstrapping method [11] is further exploited to give a confidence interval of estimation. Compared with others, our method is simpler, more systematic and more accurate. Experimental results show that the average error of our method is less than $2m$. Hence, it can better support LBAC applications.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms

Design, Security

Keywords

Statistical indoor localization, wireless signal strength, location-based access control

1. INTRODUCTION

Traditional access control systems identify and authenticate users based on something they know (e.g., password or passphrase), something they have (e.g., access token or crypto-card), or something they are (e.g., fingerprint or voice). However, none of these methods is perfect. Password may be guessed. Token can be stolen, and fingerprint is vulnerable to replay. Fortunately, the information of user location offers a new dimension for authentication and access control. For example, to grant an access to some service, we can require that a user be present at a specific location (e.g., in a room or office). Otherwise, the access is denied. It is called Location-based Access Control (LBAC) [1, 5, 6], which provides more reliable access control when combined with traditional methods. In addition, it offers the ability to trace an intruder back to a physical location if some intrusion has been detected.

Location awareness is critical for supporting location-based access control. The challenge is how to determine locations accurately and efficiently, especially in indoor environments. Existing indoor approaches utilize different types of signals such as infrared [20], ultrasound [10], and radio frequency (RF) [2, 21, 13, 15, 17, 18] to estimate locations. Among these approaches, the RF-based ones are the most promising, because they can be easily integrated with existing and widely spread 802.11 infrastructure. RF-based approaches can be further classified depending on the metrics they measure, such as Angle of Arrival (AoA) [15], Time of Arrival (ToA) [17], Time Difference of Arrival (TDoA) [18] and Received Single Strength (RSS) [2, 21, 13]. Since measuring AoA, ToA and TDoA requires special hardware such as directional antenna or fine-grained timer, localization based on WLAN signal strength seems more attractive and becomes more popular, which is also our focus.

Indoor localization approaches using wireless LAN signal strength typically consist of two phases such as offline training phase and online localization phase. In the offline phase, the signal strength from (or received by) different access points at different locations is recorded and used to build a radio map. Then, in the online phase, the measured signal strength is compared with that stored in the radio map to find the best matched and hence determine the corresponding location. RADAR [2] measures signal strength by averaging a number of samples received by several access points from a mobile client within a period of time. Horus [21] identifies different causes for signal strength variations, and proposes corresponding solutions to these variations. These solution complicates Horus system and should be adjusted for each specific indoor environment. Lim et al. [13] observed that the indoor environments are time-variant. They proposed to utilize real-time measurements for addressing environment dynamics and hence the offline

phase is no longer needed in their approach. However, they adopted a simple linear mapping between a signal strength and the logarithm of a distance, which becomes the main source of localization errors.

In this paper, we propose a statistical indoor localization method using WLAN signal strength for supporting location-based access control. In an offline phase, we fit a LOESS [3, 4, 16] local regression model on a training set to build a radio map, which stores the distribution of signal strength. In an online phase, we determine the location using Maximum Likelihood Estimation (MLE) [7, 8, 9] based on the measured signal strength and the distribution stored in the radio map. We further exploit a Bootstrapping method [11] to give the confidence interval for our estimation. Compared with existing approaches, our method is simpler and more accurate. It is systematic and can be generalized for any indoor environments. Experimental results show that the average error of our method is less than $2m$, so it can better support LBAC applications.

The paper is organized as follows: Related work is introduced in section 2. In Section 3, we discuss each component of our method including LOESS model, MLE and Bootstrapping modules in detail. Then, we present experimental results in section 4, and conclude in section 5.

2. RELATED WORK

RADAR [2] is a two-phase indoor localization system using WLAN signal strength. In the offline phase, three base stations measure the average signal strength from a mobile client and build a radio map recording locations, signal strength, and users' directions. In the online phase, it uses a K-nearest approach search a location in the radio map, which best matches the measured signal strength. RADAR has high location errors, because the simple average value can not represent the variation of signal strength precisely. Moreover, RADAR cannot be accommodated to different mobile clients whose signal strength is different.

Horus [21] is also a two-phase localization system. Unlike RADAR, Horus stores in the radio map the distribution of signal strength collected by a mobile client from different access points and determines the location using Bayes' theorem. Horus identifies different causes of signal strength variation and proposes corresponding solutions. For example, it uses an autoregressive model to handle the correlation between different samples from the same access point and utilizes a Perturbation technique to deal with small-scale variation of signal strength. To obtain a continuous location estimation, Horus exploits a time-average window to smooth the resulting location. It achieves a high accuracy, but it is too complicated and has many parameters that should be adjusted for different indoor environments. So, it is not systematic for general LBAC applications.

Lim et al. observed that the indoor environments are time-variant. which means that the environmental model learned in the offline phase may not be suitable for the data collected in the online phase. Thus, they proposed a Zero-Configuration system [13] which updates the environmental model continuously without an offline phase. However, this system assumes a simple linear relationship between a signal strength value and a distance. This assumption can not capture the dynamic property of indoor environments accurately and become the main source of location error.

3. OUR STATISTICAL METHOD

3.1 Framework of location-based access control

We consider a simple application of location-based access control: Alice is an employee of financial department of some company. She is allowed to connect to the company's server from her wireless laptop and manage a database that contains the salary information of all employees of the company. To gain access to the database (or the server), Alice must provide not only her password, but also her location information. The company's security policy requires that she be present in a particular office when managing the database. There are two reasons for enforcing this policy. The first one is for security. Bob, an attacker who steals Alice's password, may also connect to the server even from the parking lot outside the company's building. However, he could not enter into the office due to the lack of key. The second reason is for privacy. An employee's salary information is private and should not be viewed by others except for Alice. If no employee other than Alice can enter the office, the employees' privacy is fully guaranteed.

Suppose that several access points have been equipped in the company's building. we further assume that a program has been downloaded from the company's server to Alice's laptop. This program can automatically measure the signal strength from the access points and send this measured information to the server, when Alice logs on. The problem becomes how we could design a method to determine Alice's location accurately based on the measured signal strength.

3.2 Overview of our method

We propose a statistical method to determine locations based on the signal strength of access points for supporting location-based access control. Figure 1 illustrates the structure of our method that consists of an offline and an online phases.

In the offline phase, we first measure the signal strength received from different access points at each known location. Then, the offline measurements are fitted into a LOESS local regression module, which builds a radio map containing the distribution of signal strength received at each location. In the online phase, the signal strength measured at an unknown location x is processed by a Maximum Likelihood Estimator (MLE) module based on the distribution obtained in the offline phase to generate an estimation of location x. Meanwhile, a Bootstrapping module outputs a 95% confidence interval for the estimated location.

Compared with existing solutions, our method has several advantages: (1) The LOESS module produces a model independent of any physical model. So, our method does not need to study the complicated theoretical model of signal strength in indoor environments. (2) The MLE module has the nice properties such as asymptotic normality and asymptotic unbiased minimum variance estimation. So, our estimation is theoretically robust, unlike Horus that needs to be adjusted for each different environment. (3) The Bootstrapping module provides a confidence interval for location estimation, which is more meaningful than just a single estimation value. (4) Our method is simpler and more efficient, which has a higher localization accuracy.

In the rest of the section, we discuss the LOESS, MLE and Bootstrapping modules in detail.
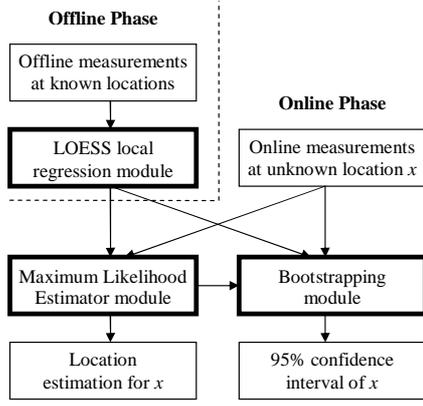
Figure 1: The structure of our statistical method

3.3 LOESS local regression module

3.3.1 Introduction to LOESS

The local regression fitting method (LOESS) [3, 4, 16] was first proposed by Cleveland in 1979. It fits curves and surfaces to noisy data with locally weighted polynomial regression. A low degree polynomial is fitted to each point in the data set by giving more weight to nearby points and less to points farther away. The biggest advantage of LOESS is that it does not need to fit a specific function to all the samples. In addition, as [16] points out, its flexibility and simplicity make it ideal for modeling very complex situations, when no clear theoretical model exists. This is the exact situation for the signal strength distribution inside a building.

3.3.2 Detailed procedure of LOESS

Given a data set of n points $\{x_1, y_1\}, \cdots, \{x_n, y_n\}$, the purpose of LOESS is to find a proper polynomial regression function g_i for each point $\{x_i, y_i\}$ such that

$$y_i = g_i(x_i) + \epsilon_i \ , \tag{1}$$

where ϵ_i is the regression error. (Note: In our localization method, x_i represents a location and y_i denotes the signal strength of some access point received at x_i.) The degree of g_i is determined by a pre-defined parameter d. When $d = 1$, g_i is a linear function. When $d = 2$, g_i corresponds to a quadratic model, that is,

$$y_i = \beta_{i,0} + \beta_{i,1} x_i + \beta_{i,2} x_i^2 + \epsilon_i \ , \tag{2}$$

where $\beta_{i,0}$, $\beta_{i,1}$ and $\beta_{i,2}$ are determined by LOESS.

As its name, *local regression*, suggests, LOESS fits the regression function to each point $\{x_i, y_i\}$ using k ($k = nq$) points that are closest to $\{x_i, y_i\}$, where q is a smoothing parameter. Let $\{x_{min}, y_{min}\}, \cdots, \{x_i, y_i\}, \cdots, \{x_{max}, y_{max}\}$ denote these k closest points. Typically, we set

$$min = i - \lfloor \frac{k-1}{2} \rfloor \ \ \text{and} \ \ max = i + \lfloor \frac{k-1}{2} \rfloor \ . \tag{3}$$

For example, $min = 2$ and $max = 7$, when $i = 5$ and $k = 6$, which means that LOESS chooses points $\{x_2, y_2\}, \cdots, \{x_7, y_7\}$ to fit regression function g_5 to point $\{x_5, y_5\}$. Equation (3) is not applicable to the case that $min < 1$ or $max > n$, but we can easily find that the k nearest points should be $\{x_1, y_1\}, \cdots, \{x_k, y_k\}$, or $\{x_{n-k+1}, y_{n-k+1}\}, \cdots, \{x_n, y_n\}$.

LOESS does not treat each of k nearest points equally. In fact, each point is assigned a weight depending on its distance to $\{x_i, y_i\}$. Let $d_{max} = \max \left(|x_j - x_i| \right)$ denote the maximum distance between x_i and x_j, for $x_j \in [x_{min}, x_{max}]$. The weight for the point at x_j is

$$w(x_j) = \left(1 - (\frac{|x_j - x_i|}{d_{max}})^3 \right)^3 \ . \tag{4}$$

Considering the polynomial model shown in equation (2) and a weighted least-squares estimator, LOESS needs to estimate $\{\hat{\beta}_{i,0}, \hat{\beta}_{i,1}, \hat{\beta}_{i,2}\}$ that minimize the following quantity:

$$Q = \sum_{j=min}^{max} w(x_j) \left(y_j - (\beta_{i,0} + \beta_{i,1} x_j + \beta_{i,2} x_j^2) \right)^2 \ . \tag{5}$$

The corresponding minimization criteria are,

$$\frac{\partial Q}{\partial \beta_{i,j}} = 0 \ \ \ \ \text{for} \ \ \ j = 0, 1, 2. \tag{6}$$

This estimation can also be expressed in matrix. Let us define

$$Y = \begin{pmatrix} y_{min} \\ \vdots \\ y_{max} \end{pmatrix}, \ \ \ \ X = \begin{pmatrix} 1 & x_{min} & x_{min}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{max} & x_{max}^2 \end{pmatrix},$$

$$\vec{\beta} = \begin{pmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \beta_{i,2} \end{pmatrix}, W = \begin{pmatrix} w(x_{min}) & & \\ & \ddots & \\ & & w(x_{max}) \end{pmatrix}.$$

The weighted least-squares estimator of $\vec{\beta}$ is:

$$\hat{\vec{\beta}} = (X^T W X)^{-1} X^T W Y \tag{7}$$

It can shown that the results calculated from equations (6) and (7) are equivalent. Given the estimation of $\vec{\beta}$ as

$$\hat{\vec{\beta}} = \begin{pmatrix} \hat{\beta_{i,0}} \\ \hat{\beta_{i,1}} \\ \hat{\beta_{i,2}} \end{pmatrix},$$

LOESS fits a quadratic model at point $\{x_i, y_i\}$ as

$$\hat{y}_i = g_i(x_i) = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} x_i + \hat{\beta}_{i,2} x_i^2 \ . \tag{8}$$

The regression function $g_i(x_i)$ is calculated repeatedly at every point in the data set $\{x_1, y_1\}, \cdots, \{x_n, y_n\}$.

3.3.3 Choosing the smoothing parameter q

The most important two parameters controlling LOESS are d and q. Once we determine the value of d (e.g., $d = 2$), q is chosen from $[(d + 1)/n, 1]$, which controls how much amount of data is used in each polynomial regression. However, which value of q is the best?

In our method, we determine the value of q by finding the model minimizing Akaike's Information Criterion (AIC) [12]. AIC is one of the most commonly used penalized model selection criteria. One version of the bias-corrected AIC value for a LOESS model might be

$$AIC_C = log(\hat{\sigma}^2) + 1 + \frac{2(Trace(L) + 1)}{n - Trace(L) - 2} \ , \tag{9}$$

where n is the number of data points and $\hat{\sigma}$ is the standard error of data. $Trace(L)$ is the trace of matrix L, which is the

smoothing matrix of the LOESS model. L defines the linear relationship between the fitted and observed dependent variable values. That is, L satisfies

$$\hat{Y} = LY \ , \tag{10}$$

where $\hat{Y}$ can be calculated using equation (8).

3.3.4 Fitting LOESS model on training set

Considering a system with m access points and n known test locations, we first collect the signal strength from all access points at each test location in an offline training phase. Our training set is $\{x_1, s_{1,j}\}, \cdots, \{x_i, s_{i,j}\}, \cdots, \{x_n, s_{n,j}\}$, where x_i for $i = 1, \cdots, n$ denotes a test location and $s_{i,j}$ for $j = 1, \cdots, m$ denotes the signal strength of the j-th access point received at location x_i. The LOESS model on the training set can be expressed as

$$s_{i,j} = g_{i,j}(x_i) + \epsilon_{i,j} \ , \tag{11}$$

where $g_{i,j}$ and $\epsilon_{i,j}$ denote the regression function and regression error for the j-th access point at location x_i.

Then, we estimate $g_{i,j}$ based on the LOESS model and the chosen q. We assume that the regression error $\epsilon_{i,j}$ satisfies some normal distribution, that is,

$$\epsilon_{i,j} \sim \mathcal{N}(0, \ \sigma_{i,j}^2) \ , \tag{12}$$

where $\sigma_{i,j}^2$ denotes some variance. Thus, the signal strength also satisfies a normal distribution, that is,

$$s_{i,j} \sim \mathcal{N}(g_{i,j}(x_i), \ \sigma_{i,j}^2) \ . \tag{13}$$

This distribution is stored in the server and will be used in MLE module to determine locations.

3.4 Maximum likelihood estimator module

3.4.1 Introduction to MLE

Maximum Likelihood Estimator (MLE) [7, 8, 9] was formally proposed by Fisher. Given a large size of samples, MLE provides an unbiased minimum variance estimation and its estimates are approximately normally distributed.

Considering a probability distribution family with a probability density (mass) function f_θ, which is parameterized by unknown θ (a scalar or a vector). Given a set of observations $\{x_1, x_2, ..., x_k\}$ drawn from the distribution f_θ, the likelihood function for this set of observations is

$$L(\theta) = f_\theta(x_1, x_2, ..., x_n | \theta) \ . \tag{14}$$

If the observations are independent of each other, the likelihood function can be further written as

$$L(\theta) = \prod_{i=1}^{k} f_\theta(x_i | \theta) \ . \tag{15}$$

Applying logarithmic transformation to both sides of equation (15), the likelihood can also be expressed as

$$l(\theta) = \log\big(L(\theta)\big) = \sum_{i=1}^{k} \log\big(f_\theta(x_i | \theta)\big) \ . \tag{16}$$

The maximum likelihood estimator for θ, denoted as $\hat{\theta}$, is the value that maximizes the likelihood $L(\theta)$ or $l(\theta)$.

3.4.2 Location estimation using MLE

In the online phase, a user collects the signal strength $s_1, \cdots, s_j, \cdots, s_m$ from m access points at an unknown location x. The user sends the measured signal strength to the server, which then estimates location x using our MLE module and decides if an access should be granted to the user.

From the LOESS fitting result and the normal distribution in equation (13), the probability density function $f_j(s_j | x)$ for the signal strength s_j of the j-th access point received at location x is

$$f_j(s_j | x) = \phi_{\mathcal{N}}\big(\frac{s_j - g_j(x)}{\sigma_j}\big) \ , \tag{17}$$

where g_j and σ_j denote the regression function and the corresponding standard deviation of the j-th access point. $\phi_{\mathcal{N}}$ is the standard normal density function such that

$$\phi_{\mathcal{N}}(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \ . \tag{18}$$

The likelihood function for the set of signal strength received from all of m access points at location x is

$$L(x) = \prod_{j=1}^{m} f_j(s_j | x) = \prod_{j=1}^{m} \phi_{\mathcal{N}}\big(\frac{s_j - g_j(x)}{\sigma_j}\big) \ . \tag{19}$$

In practice, we often calculate the logarithmical likelihood

$$l(x) = \log\big(L(x)\big) = \sum_{j=1}^{m} \log\big(\phi_{\mathcal{N}}(\frac{s_j - g_j(x)}{\sigma_j})\big) \ . \tag{20}$$

In our method, the server estimates location x using maximum likelihood estimator $\hat{x}$, which maximizes $L(x)$ or $l(x)$.

3.5 Bootstrapping module

Bootstrapping [11] is a statistical method for estimating the distribution of an estimator by re-sampling the original data. With this method, we can easily compute the confidence interval of estimation with higher accuracy than with other methods based on normal-approximation. In the online phase, we take advantage of a Bootstrapping module to give the confidence interval of our estimation, while none of existing approaches can report such a confidence interval.

The procedure of Bootstrapping module consists of two steps. In step 1, the signal strength is re-sampled from the distribution $s_j(x) \sim \mathcal{N}(g_j(x), \sigma_j^2)$, which are calculated in the offline phase by LOESS module, to get a new sampled training set. In step 2, a new estimator for x is calculated using MLE from the new training set. Iterating these two (i.e., re-sampling and estimating) steps usually several thousands times, we get a sample distribution for the new estimator $\hat{x}$. A 95% confidence interval is then calculated by selecting 2.5% and 97.5% quantile values as the lower and upper bounds.

4. EXPERIMENTAL RESULTS

4.1 Experiment setup

Our method is evaluated by experiments. Signal strength is obtained from three access points: (1) Netgear WGR614, (2) Linksys WRTSL54GS, and (3) IASTATE hot spot, which are deployed on the third floor of Coover Hall at Iowa State
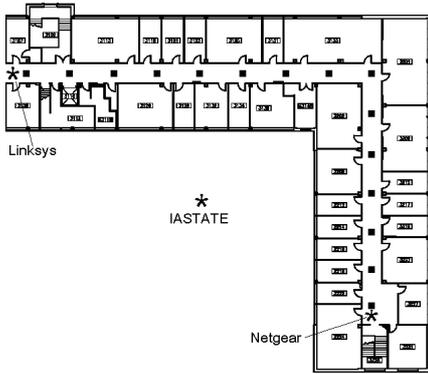
Figure 2: The floor plan of Coover Hall for experiments, where the black blocks represent test locations and the stars are access points

University. Figure 2 shows the plan of the floor, which contains a $27.1m$ north-south corridor and a $30.8m$ east-west corridor. The Linksys and Netgear are placed at the end of corridors, while the IASTATE access point belongs to Iowa State University and is out of our control. 48 test locations along the middle line of corridors are selected with 1.23 meters between each pair of neighbors. Among these 48 locations, half belongs to the training set, while the other half forms a location set used for location estimation. In either set, the locations are selected from the total 48 ones by every other one.

A Dell Inspiron 8200 laptop equipped with Windows XP, NetStumbler 0.4.0 [14], and a Linksys WUSB54GP external wireless network card is used to collect the signal strength at all of 48 test locations. At each location, the signal strength from each access point is measured every 0.5 second within 1 minute (120 samples from each access point). The median of these 120 samples is used for final analysis. Figure 3 plots the median of signal strength of each access point at every test location, which is identified by the distance between the test location and a fixed start point.

4.2 LOESS local regression on training set

We utilize the statistical computing package R [19] to perform LOESS local regression on the training set. The degree of regression function is $d = 2$ and the value of smoothing parameter q is determined based on AIC metric. Figure 4 shows the results of LOESS regression for each access points with 95% confidence interval, while we can see that the LOESS model smoothly matches the real data shown in Figure 3.

4.3 Maximum likelihood estimation on location set

We apply MLE to the location set and estimate the location of each point in the set. Localization accuracy is evaluated by the difference between the estimated locations and the true ones. Figure 5 plots the likelihood function of signal strength received at location $x = 18.486m$. The estimated location is $18.251m$ that is the peak position of MLE, so the estimation error is $0.235m$.

Figure 6 shows the experimental cumulative distribution function (CDF) of MLE error. Table 1 compares the location errors among different methods with 25%, 50% and 75%

Table 1: Comparison of location error

	25%	50%	75%
RADAR	$1.92m$	$2.94m$	$4.69m$
Lim's method	$0.97m$	$2.57m$	$3.56m$
Our method	$0.69m$	$1.71m$	$2.53m$

of total estimated locations. It is clear that our method outperforms RADAR [2] and Lim's approach [13] in terms of localization accuracy.

4.4 Impact of the size of training set

Increasing the size of training set, i.e., choosing more test locations, can reduce location error, because the LOESS model has more information about the indoor environment. We study the impact of the size of training set by adjusting the interval between neighbor test locations. Figure 7 plots the mean of MLE error as a function of the interval. It shows that increasing the interval from $1m$ to $6\ m$ only raises the MLE error by 25% (from $2m$ to $2.5m$), meanwhile, we save 83% of the time for collecting the training data. So, our method is robust to the changes of the size of training set.

4.5 Impact of the number of access points

Increasing the number of access points should make estimation more accurate, because more information can be exploited to determine locations. Figure 8 shows the impact of increasing the number of access points, where the mean of MLE error is reduced by 20% (from $2m$ to $1.6m$) as the number of access points doubled (form 3 to 6).

5. CONCLUSIONS AND FUTURE WORK

We propose a statistical indoor localization method for supporting location-based access control (LBAC). Our method uses LOESS regression to build the distribution of WLAN signal strength and estimates locations use MLE method. It can also give a 95% confidence interval for its estimation with the help of a Bootstrapping module. Compared with others, our method is simpler and has a higher accuracy. It does not need to study any physical model of indoor signal strength and can provide a meaningful confidence interval. Experimental results show that the location error of our method is less than $2m$ and hence it can better support LBAC applications.

Our method can be applied to the locations of two or three dimensions almost without changes. We plan to conduct extensive experiments on two and three dimensional estimation in the future. We will also study how to utilize real-time information update to our LOESS model.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] C. Ardagna, M. Cremonini, E. Damiani, S. Vimercati, and P. Samarati, "Supporting Location Based Access Control Policies", in *Proc. 2006 ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pp. 212–222, 2006.

(a) Netgear   (b) Linksys   (c) IASTATE
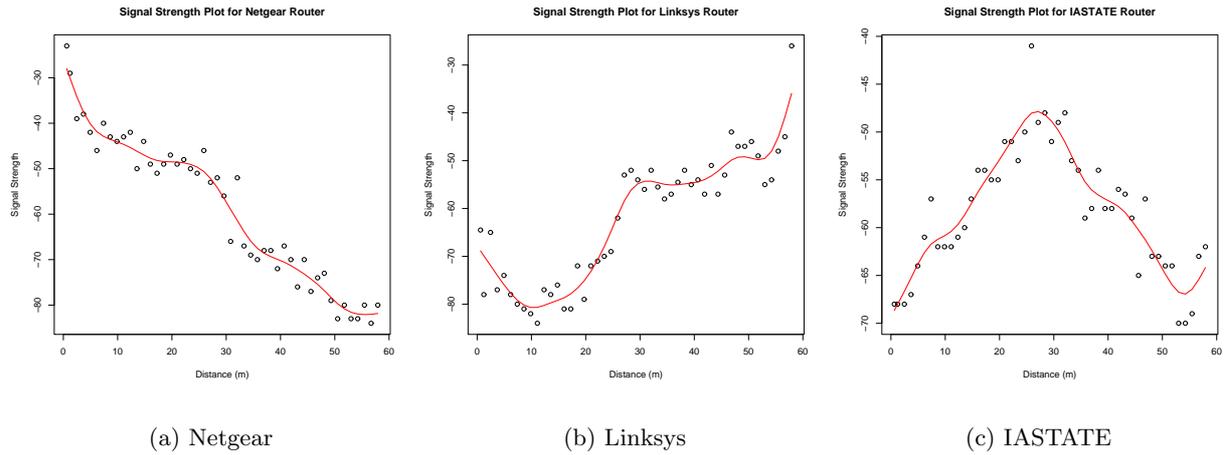
Figure 3: The signal strength of different access points as a function of test locations

(a) Netgear   (b) Linksys   (c) IASTATE
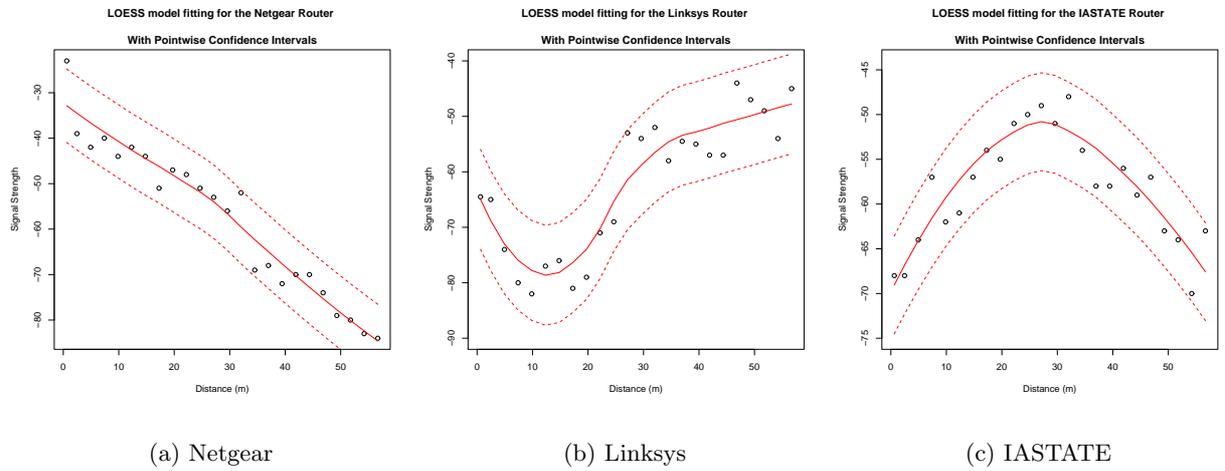
Figure 4: LOESS regression results for different access points with 95% confidence interval

[2] P. Bahl and V. Padmanabhan, "RADAR: An In-building RF-based User Location and Tracking System", in *Proc. IEEE INFOCOM*, pp. 775–784, 2000.

[3] W. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots", in *Journal of the American Statistical Association*, Vol. 74, pp. 829–836, 1979.

[4] W. Cleveland and S. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", in *Journal of the American Statistical Association*, Vol. 83, pp. 596–610, 1998.

[5] M. Covington, W. Long, S. Srinivasan, A. Dey, M. Ahamad, and G. Abowd, "Securing Context-Aware Applications Using Environment Roles", in *Proc. 6th ACM Symposium on Access Control Models and Technologies (SACMAT)*, 2001.

[6] D. Denning and P. F. MacDoran, "Location-Based Authentication: Grounding Cyberspace for Better Security", in *Computer Fraud and Security*, 1996.

[7] R. Fisher, "On an Absolute Criterion for Fitting Frequency Curves", in *Statistical Science*, Vol. 12, No. 1, pp. 39–41, 1997.

[8] R. Fisher, "On the Mathematical Foundations of Theoretical Statistics", in *Philosophical Transactions of the Royal Society of London, Series A*, Vol. 222, pp. 309–368, 1992.

[9] R. Fisher, "The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients", in *Journal of the Royal Statistical Society*, Vol. 85, No. 4, pp. 597–612, 1992.

[10] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The Anatomy of a Context-aware Application", in *Proc. ACM MobiCom*, 1999.

[11] P. Hall, "The Bootstrap and Edgeworth Expansion", Springer, 1992, ISBN: 0387945083.

[12] C. Hurvich, J. Simonoff, and C. Tsai, "Smoothing Parameter Selection in Nonparametric Regression
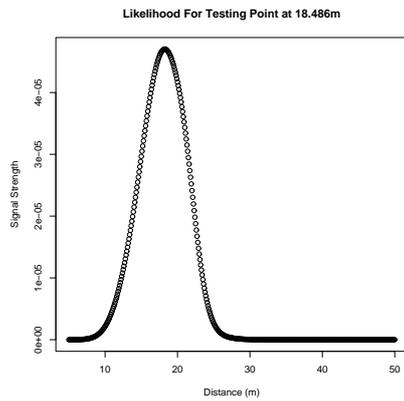
**Figure 5: The likelihood function for test location at 18.486m**
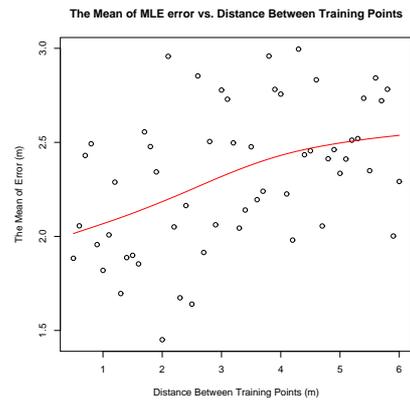
**Figure 7: The mean of MLE error as a function of the interval between test locations in training set**
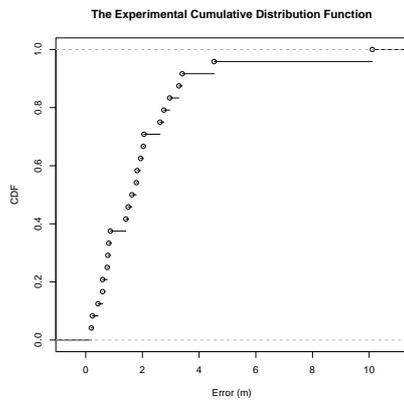
**Figure 6: The Cumulative Distribution Function (CDF) of MLE error**
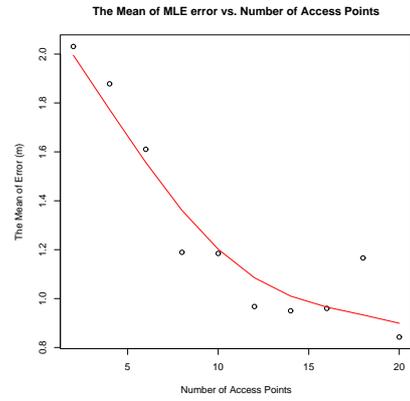
Figure 8: The mean of MLE error as a function of the number of access points

Using an Improved Akaike Information Criterion", in *Journal of the Royal Statistical Society, Series B*, Vol. 60, No. 2, pp. 271–293, 1998.

[13] H. Lim, L. Kung, J. Hou, and H. Luo, "Zero-Configuration, Robust Indoor Localization: Theory and Experimentation", in *Proc. IEEE INFOCOM*, 2006.

[14] NetStumbler, http://www.netstumbler.com

[15] D. Niculescu and B. Nath, "VOR Base Stations for Indoor 802.11 Positioning", in *Proc. ACM MobiCom*, 2004.

[16] NIST, *Engineering Statistics Handbook*, http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm

[17] Pinpoint, http://www.pinpoint.com/

[18] N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-support System", in *Proc. ACM MobiCom*, 2000.

[19] The R Project for Statistical Computing, http://www.r-project.org/

[20] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The Active Badge Location System", in *ACM Transactions on Information Systems*, Vol. 10, No. 1, pp. 91–102, 1992.

[21] M. Youssef, and A. Agrawala, "The Horus WLAN Location Determination System", in *Proc. ACM MobiSys*, pp. 205–218, 2005.