# Signal Strength Based Indoor Geolocation[*]

## Yongguang Chen and Hisashi Kobayashi
Department of Electrical Engineering
School of Engineering and Applied Science
Princeton University
Princeton, NJ 08544

*Abstract* - **We have investigated the indoor geolocation based on signal strength modeling. Linear, compensated linear, and multiple regression methods have been applied to set up signal strength models by using simulated data. We have also analyzed this modeling method to better understand the relation between the location error and the signal strength error. Some important results have been obtained to help us determine proper placement of Access Points (APs) and evaluate the range of location error. A simulation experiment has been conducted based on typical parameters of IEEE802.11b MAC.**

*Index terms* - **WLAN, geolocation, signal strength, IEEE802.11b.**

## I. INTRODUCTION

With the development of WLANs (Wireless Local Area Networks), there is an increasing level of interest in developing the technology to "geolocate" WLAN users, especially in an indoor environment. Positioning and tracking of an indoor user based on radio signals will encounter a considerable degree of technical difficulty because various objects such as floors, walls and human bodies within a confined space will contribute to a rather complex form of attenuation and fading of the radio signals to be used for geolocation. [1] A majority of wireless geolocation techniques are based on such information as TOA (the time of arrival), TDOA (the time difference of arrival), and DOA (the direction of arrival). But geolocation based on these techniques is reliable only when line-of-sight signals are dominant, hence it will not be applicable to an indoor environment. Furthermore, a TOA or TDOA based approach requires accurate synchronization between transmitters and receivers.

We therefore explore an alternative geolocation method, that is, a signal strength based approach. Instead of measuring the time or angle of signal arrival, the signal strength method makes use of the level of signal power (or energy) sensed by an MS (mobile station) regarding the signals transmitted by reference base stations or APs (access points in the IEEE802.11 terminology). This signal strength based approach may be also possible in a reversed situation, where the signal from an MS is sensed by multiple APs. This second approach would relieve an individual MS from the task of computing its position or processing and transferring relevant information to some BS (base station) or AP, as would be required in the first approach. However, a set of signals from different MSs must be designed in such a manner that APs can distinguish the signals from different MSs.

As early as in the 1960's, the signal attenuation model has been proposed as an approach to locate vehicles in motion on the street. [2] Nevertheless, the signal strength based geolocation is still an unexplored technique for locating WLAN users in an indoor environment. In our research we use, as our starting point, a recent work reported by Bahl and Padmanabhan. [3] Before reviewing their analysis, we first introduce a simple signal propagation model which is based on a signal predictor variable, where the observed variable is the signal strength (in *dBm*), and the predictor (or controlled) variable is the distance from a reference position (also in logarithm), and the main parameter to be estimated (i.e., regression coefficient) is the exponent value $\alpha$ that determines path loss of the signal when the distance from the signal source is given. Then we use Bahl's empirical signal propagation model which is also based on a linear regression analysis and in which the observed variable of signal strength has been compensated for the attenuation caused by the walls intervening between the MS and AP before applied to the regression analysis.

We then extend this linear regression model to a multiple regression model by adding another predictor variable, i.e., the wall attenuation factor, denoted *WAF* [*dB*]. There may exist walls intervening between a possible MS location (to be estimated in geolocation) and a given reference position. We then evaluate the improvement of this multiple regression model over the linear regression model by comparing their coefficients of determination and standard deviations. In order to carry out this statistical analysis, we resort to a simulation technique.

In the last section, we discuss how the statistical model for signal propagation developed in the preceding section can be utilized to assess the error in geolocating an MS by this signal strength based method. The magnitude of location error is a function of the position of an MS and the set of AP positions. Hence this error analysis can be further used to determine optimal locations of the set of APs.

## II. RADIO PROPAGATION MODELING

### A. A Simple Signal Propagation Model

If we can assume that the signal strength is related only to the distance between the transmitter and the receiver, the following simple signal propagation model may hold.

$$P(r)[dBm] = P(r_0)[dBm] - 10\alpha \log(r/r_0), \qquad (1)$$

---

where $P(r)$ is the power received by a given MS whose distance from a given transmitter or AP is $r$ (meters); $r_0$, the reference distance from the transmitter; and $P(r_0)$, the signal power at this reference point. The parameter $\alpha$, called the exponent value, indicates the rate at which the path loss increases as distance $r$ increases.

This model does not take into account walls that may exist in a building. Therefore, if we rely on this simple model to calculate the signal strength received at a given indoor position, the signal strength may be significantly overestimated, especially in a building with many rooms isolated by walls.

### B. An Empirical Signal Propagation Model

An empirical model of signal strength is introduced in [3].

$$P(r)[dBm] = P(r_0)[dBm] - 10\alpha \log(r/r_0) - l \cdot WAF , \quad (2)$$

where $l$ is the number of walls between the transmitter and the receiver; and $WAF$ is the wall attenuation factor.

### C. Linear Regression Analysis

We can estimate the unknown parameters $P(r_0)$ and $\alpha$ by applying linear regression analysis. In the simple signal propagation model, the measured signal power is treated as the observed variable. In case of the empirical signal propagation model, however, we first determine $WAF$ [$dB$] by comparing the average strength of a signal subject to intervening walls with the signal strength that would be observed in a line-of-sight environment. If we modify each value of the measured signal strength by compensating for the signal loss caused by the intervening walls between the transmitter and the receiver, linear regression analysis can be used. By compensation we mean that (2) should be changed to

$$\begin{aligned}P_W(r)[dBm] &= P(r)[dBm] + l \cdot WAF \\ &= P(r_0)[dBm] - 10\alpha \log(r/r_0).\end{aligned} \quad (3)$$

Let us define

$$\rho_i = -10\log(r_i/r_0), i = 1, 2, \cdots, m , \quad (4)$$

where $m$ is the number of samples. We then have the following estimates [4]

$$\hat{\alpha} = \frac{\sum_{i=1}^{m}[\rho_i - \overline{\rho}]P_W(r_i)}{\sum_{i=1}^{m}(\rho_i - \overline{\rho})^2} , \quad (5)$$

and

$$\hat{P}(r_0)[dBm] = \overline{P}_W[dBm] - \hat{\alpha}\overline{\rho} , \quad (6)$$

where

$$\overline{\rho} = \frac{1}{m}\sum_{i=1}^{m}\rho_i , \quad (7)$$

$$\overline{P}_W[dBm] = \frac{1}{m}\sum_{i=1}^{m}P_W(r_i)[dBm] . \quad (8)$$

The *coefficient of determination*, $R^2$, represents the goodness of regression. It is defined as follows:

$$\begin{aligned}R^2 &= \frac{\text{Variance of signal strength explained by regression}}{\text{Total variation of signal strength}} \\ &= \frac{\sum_{i=1}^{m}[\hat{P}_W(r_i) - \overline{P}_W]^2}{\sum_{i=1}^{m}[P_W(r_i) - \overline{P}_W]^2} .\end{aligned} \quad (9)$$

The *standard deviation* of the predicted signal strength is given by

$$\sigma_P = \sqrt{\frac{1}{m}\sum_{i=1}^{m}[P_W(r_i) - \hat{P}_W(r_i)]^2} . \quad (10)$$

### D. Multiple Regression Analysis

Multiple regression is an extension of the simple regression analysis to take into account effects of more than one predictor variable on the dependent variable. As shown in (2), we now intend to determine the three factors (i.e., $P(r_0)$, $\alpha$, and $WAF$) simultaneously using the two predictor variables (i.e., the distance $r$, and the number of walls, $l$). Assuming that $m$ independent samples are taken to construct this multiple regression model, we now have, instead of (2), the following matrix equation.

$$\mathbf{P} = \mathbf{G}\boldsymbol{\beta} , \quad (11)$$

where

$$\mathbf{P} = \begin{bmatrix} P(r_1) \\ P(r_2) \\ \vdots \\ P(r_m) \end{bmatrix}, \mathbf{G} = \begin{bmatrix} 1 & -10\log(r_1/r_0) & -l_1 \\ 1 & -10\log(r_2/r_0) & -l_2 \\ \vdots & \vdots & \vdots \\ 1 & -10\log(r_m/r_0) & -l_m \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} P(r_0) \\ \alpha \\ WAF \end{bmatrix}. \quad (12)$$

The least-square estimate of the unknown parameter vector $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{P}, \quad (13)$$

and the *coefficient of determination* is

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{G}'\mathbf{P} - m\overline{P}^2}{\mathbf{P}'\mathbf{P} - m\overline{P}^2} , \quad (14)$$

where

$$\overline{P}[dBm] = \frac{1}{m}\sum_{i=1}^{m}P(r_i)[dBm] . \quad (15)$$

The *standard deviation* is given by

$$\sigma_P = \sqrt{\frac{1}{m}(\mathbf{P}' - \hat{\boldsymbol{\beta}}'\mathbf{G}')\mathbf{P}} . \quad (16)$$

### E. Simulation Experiments

In order to evaluate the feasibility and effectiveness of the proposed statistical model, we conducted a simulation experiment as our first step. The system parameters designed in the simulation are as follows:

$P(r_0) = 160W = 52dBm$, and $\alpha = 2$.

As for the range of distance we assumed 70 distinct values with the minimum at $1m$ and the maximum at $35.5m$. We assume that there is a wall at every $7m$.

As for the *WAF* we assumed that percentage of the signal energy that goes through a wall is uniformly distributed between 30% and 70%. This translates to the loss factor *WAF* [*dB*] to be between $-10\log0.3 = 5.23dB$ and $-10\log0.7 = 1.55dB$.

In addition we assume the aberration of power $P(r)$ [*dB*] is subject to random error of Gaussian distribution with zero mean and $2.5dB$ standard deviation.

The estimate of factors, goodness of regression, and the standard deviation for the three regression methods are compared in TABLE I. From the results we find that the multiple regression model can improve the goodness of regression and decrease the deviation of predicted values.

TABLE I
COMPARISON OF DIFFERENT REGRESSION MODELS

| Regression Analysis | Estimate of factors | | | $R^2$ | $\sigma_P$ |
| --- | --- | --- | --- | --- | --- |
| | $P(r_0)(dBm)$ | $\alpha$ | $WAF(dB)$ | | |
| *Linear* | 59.84 | 3.26 | - | 92.13% | 3.39 |
| *Compensated Linear* | 53.52 | 2.15 | - | 88.63% | 2.74 |
| *Multiple* | 53.36 | 2.12 | 3.09 | 94.85% | 2.74 |

Comparison of the simulated signal strength data versus the predicted one is shown in Fig. 1. In multiple regression, the curve is linear only within those MS positions encountering the same number of walls intervening between the MS and the AP. When the number of walls increases by one, the signal will drop its power level and remain linear till encountering the next wall.
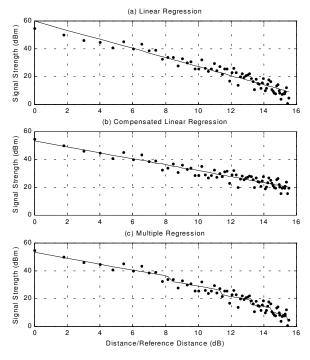


Fig. 1. Simulated data vs. predicted signal strength

## III. ESTIMATION OF LOCATION ERROR

### A. Analysis Methodology

We now investigate the relation between the location error and signal strength error. By applying a differential operation to both sides of (2) with respect to two coordinates $x$ and $y$, we have:

$$dP_i(x, y) = -\frac{10\alpha_i}{\ln 10}(\frac{x - x_i}{r_i^2}dx + \frac{y - y_i}{r_i^2}dy),\ i = 1, 2, \cdots, N\ , (17)$$

where $P_i(x, y)$ is the strength of the signal from $AP_i$ received at $(x, y)$; $(x_i, y_i)$, the coordinate of $AP_i$; $\alpha_i$, the exponent value of path loss for the signal coming from $AP_i$; $r_i$, the distance between MS and $AP_i$, i.e., $r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ ; and $N$, the number of APs.

The set of (17) can be written in matrix form as

$$d\mathbf{P} = \mathbf{H} \cdot d\mathbf{r}\ , (18)$$

where

$$d\mathbf{P} = \begin{bmatrix} dP_1 \\ dP_2 \\ \vdots \\ dP_N \end{bmatrix}, d\mathbf{r} = \begin{bmatrix} dx \\ dy \end{bmatrix}, \mathbf{H} = \begin{bmatrix} -\frac{10\alpha_1}{\ln 10}\frac{x - x_1}{r_1^2} & -\frac{10\alpha_1}{\ln 10}\frac{y - y_1}{r_1^2} \\ -\frac{10\alpha_2}{\ln 10}\frac{x - x_2}{r_2^2} & -\frac{10\alpha_2}{\ln 10}\frac{y - y_2}{r_2^2} \\ \vdots & \vdots \\ -\frac{10\alpha_N}{\ln 10}\frac{x - x_N}{r_N^2} & -\frac{10\alpha_N}{\ln 10}\frac{y - y_N}{r_N^2} \end{bmatrix} . (19)$$

Then by applying the method of least-square estimation, [5, 6] we find that an error in power estimation, $d\mathbf{P}$, will result in the following estimation of location error:

$$d\hat{\mathbf{r}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'d\mathbf{P}\ . (20)$$

We assume that signal strength estimation error has zero mean and variance $\sigma_P^2$, and these errors for different AP's are independent with each other, i.e.,

$$\mathbf{E}[dP_i] = 0, \mathrm{cov}(dP_i, dP_j) = \begin{cases} \sigma_P^2, & i = j \\ 0, & i \neq j \end{cases}\ i, j = 1, 2, \cdots, N\ , (21)$$

where $\sigma_P$ is the standard deviation of predicted signal strength derived from (10) and (16). Then the covariance matrix of the error estimate $d\hat{\mathbf{r}}$ is given by

$$\mathrm{cov}(d\hat{\mathbf{r}}) = \sigma_P^2(\mathbf{H}'\mathbf{H})^{-1} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{bmatrix}. (22)$$

The standard deviation of location error is finally estimated as

$$\sigma_r = \sqrt{\sigma_x^2 + \sigma_y^2}\ . (23)$$

### B. Geometrical Distribution of Location Error

The simulation condition for location error analysis is the same as that set for regression analysis except that a data smoothing method is adopted to decrease the variance in

simulated data when we take multiple measurements of signal power at a given sampling location.

We assume that there are five APs installed in a building with their coordinates being $AP_1$ (15$m$, 15$m$), $AP_2$ (15$m$, -15$m$), $AP_3$ (-15$m$, -15$m$), $AP_4$ (-15$m$, 15$m$), and $AP_5$ (0.1$m$, 0.1$m$). If an MS at a given location receives signals from these APs, its position can be determined by triangulation or least square estimation.
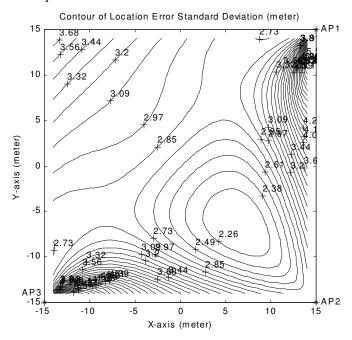


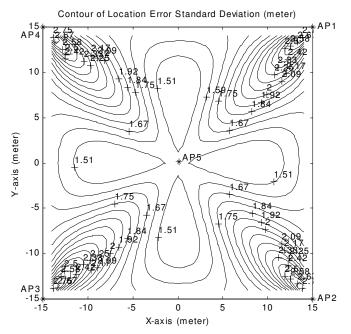Fig. 2. Geometrical distribution of location error (3 APs)



Fig. 3. Geometrical distribution of location error (5 APs)

Fig. 2, 3 shows the contours of $\sigma_r$ when three APs (i.e., $AP_1$, $AP_2$, and $AP_3$), five APs (i.e., $AP_1$ through $AP_5$) are used for geolocation, respectively. As shown in TABLE II, the location accuracy may be improved if more APs are deployed.

The location error is found to be greater along either of the baselines. The reason is that when the MS is along a baseline, the two circles (with their centers at $AP_1$ and $AP_2$) in determining the position of the MS will not intersect with each other, but touch each other at the position of the MS. Then the horizontal error $dx$ can be very large although $dy$ is small. Use of the third AP (e.g., $AP_3$) in triangulation should help decrease $dx$, but its effect is minimal in the region close to $AP_1$, because the signal from $AP_3$ is weak.

TABLE II
COMPARISON OF LOCATION ERROR

| Deployed APs | Location Error (meters) | |
| --- | --- | --- |
| | Maximum | Minimum |
| AP1, AP2 | 15.98 | 2.75 |
| AP1, AP2, AP3 | 5.64 | 2.12 |
| AP1, AP2, AP3, AP4 | 2.80 | 2.06 |
| AP1, AP2, AP3, AP4, AP5 | 2.80 | 1.41 |

## IV. CONCLUSION

From our mathematical analysis and simulation experiments, we are reasonably confident that the indoor geolocation based on signal strength should be feasible as long as we can construct a sound method for signal propagation in the indoor environment. Our analysis shows that the location error of an MS can be controlled by three factors, i.e., (1) the goodness of regression analysis, (2) the number and proper placement of APs, and (3) the geometrical relation among the positions of the MS and APs. The relatively high location accuracy obtained in our simulation and analysis is very promising. We believe this signal strength based indoor geolocation will be useful not only in a WLAN environment, but also as an augmenting technique to enhance the existing geolocation technique, especially GPS-based geolocation, which is known to be unreliable once an MS is inside a building.

REFERENCES

[1] Kaveh Pahlavan, Prashant Krishnamurthy, and Jacques Beneat, "Wideband radio propagation modeling for indoor geolocation applications," *IEEE Communications Magazine*, April 1998.
[2] William G. Figel, Neal H. Shepherd, and Walter F. Trammell, "Vehicle location by a signal attenuation method," *IEEE Trans. On Vehicular Technology*, vol. 18, no. 3, November 1969.
[3] Paramvir Bahl and Venkata N. Padmanabhan, "User location and tracking in an in-building radio network," *Technical Report*, *Microsoft Research*, February 1999.
[4] Hisashi Kobayashi, *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology.* Addison-Wesley Publishing Company, Inc., 1978.
[5] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory* (Part 1). John Wiley and Sons Inc., 1968.
[6] Charles E. Cook and Marvin Bernfeld, *Radar Signals: An Introduction to Theory and Application*, Academic Press Inc., New York, 1967.