Handling Samples Correlation in the Horus System

Moustafa Youssef and Ashok Agrawala
Department of Computer Science and UMIACS
University of Maryland
College Park, Maryland 20742
Email: {moustafa, agrawala}@cs.umd.edu
UMIACS-TR-2003-75 and CS-TR-4506
June 31, 2003

Abstract— **We present an autoregressive model for modeling samples autocorrelation from the same access point in WLAN location determination systems. Our work is in the context of the *Horus* system, which is a *probabilistic* WLAN location determination system. We show that the autocorrelation between consecutive samples from the same access point can be as high as 0.9. Using our model, we describe a technique to use multiple signal strength samples from each access point, taking the high autocorrelation into account, to achieve better accuracy. Implementation of the technique in the *Horus* system shows that the average system accuracy is increased by more than 50%. Our results show that assuming independence of samples from the same access point can lead to degraded performance as the number of samples used in the estimation algorithm is increased, due to the wrong independence assumption. We also discuss how to incorporate the new technique with other algorithms for enhancing the performance of WLAN location determination systems.**

Keywords— **System design, Experimentation with real networks/Testbeds, Statistics.**

I. Introduction

Horus is an RF-based location determination system [1]–[3]. It is currently implemented in the context of 802.11 wireless LANs [4]. The system uses the signal strength returned from the access points to infer the user location. Since the wireless cards measure the signal strength information of the received packets as part of their standard operation, this makes the *Horus* system a purely software solution on top of the wireless network infrastructure. A large class of applications, including [5] location-sensitive content delivery, direction finding, asset tracking, and emergency notification, can be built on top of the *Horus* system.

WLAN location determination is an active research area [1]–[3], [6]–[13]. 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.

Radio-map based techniques can be categorized into two broad categories: deterministic techniques and probabilistic techniques. *Deterministic techniques* [6]–[8] represent the signal strength of an access point at a location by a scalar value, for example, the mean value, and use non-probabilistic approaches to estimate the user location. For example, in the *Radar* system [6], [7] the authors use nearest neighborhood techniques to infer the user location. On the other hand, *probabilistic* techniques [1]–[3], [9]–[13] store information about the signal strength distributions from the access points in the radio map and use probabilistic techniques to estimate the user location. For example, the Nibble system [9], [10] uses a Bayesian Network approach to estimate the user location.

The *Horus* system lies in the probabilistic techniques category. Its goal is to identify the noisy characteristics of the wireless channel and to develop techniques to handle them. In this paper, we analyze one aspect of the temporal characteristics of the wireless channel: samples correlation from the same access point. We show that consecutive samples can have correlation as high as 0.9. The main challenge is how to use multiple samples to obtain better location estimate technique despite this high correlation value. Our approach is to treat the samples collected from an access point at a given location as a time series [14] and use the time-series analysis techniques to study their characteristics. More specifically, we describe an autoregressive model that captures the correlation of samples from the same access point. Based on the autoregressive model, we present a technique that uses multiple samples from each access point, to increase the accuracy of the *Horus* system. We present the result of implementing the new technique and compare its accuracy with that of the original *Horus* system. We also discuss how to incorporate the proposed technique with other techniques for enhancing the performance of WLAN location determination systems.

The rest of the paper is structured as follows: in the next section, we discuss related work. In section III we present an overview of the *Horus* system and briefly introduce autoregressive models. Section IV shows the temporal characteristic of the samples received from an access point and analyze the autocorrelation of samples. We describe our autoregressive model to capture the signal strength samples correlation and the technique that uses this model to enhance the accuracy of the *Horus* system in Section V. We present the results of implementing the new technique and compare its accuracy to the accuracy of the original technique in Section VI. Finally, Section VII discusses the main findings of the paper and provides concluding remarks.

II. Related Work

In this section, we describe other techniques that use multiple samples to enhance the performance of WLAN location determination systems. We show how the proposed technique relates to them.

A. Signal Strength Space Averaging

The authors of the *Radar* [6], [7] system, a *deterministic* location determination technique, were the first to propose using multiple signal strength samples to obtain better estimation accuracy. Their technique is to average the received samples and use the average value in the k-nearest neighborhood algorithm to determine the best location estimate. Their results indicate that using more samples in the averaging process leads to better accuracy.

The work in this paper is concerned with *probabilistic* location determination techniques in which the process of using multiple samples to obtain a location estimate is more involved. For example, if the system averages n samples, the system needs to calculate the probability of the average value using the distribution of the average of n original distribution. Obtaining this distribution is not trivial if the samples are not independent. We address this issue in Section V.

B. Physical Location Space Averaging

Different systems, e.g. [6], [7], [11]–[13], proposed to use averaging in the *physical-location space*. The system

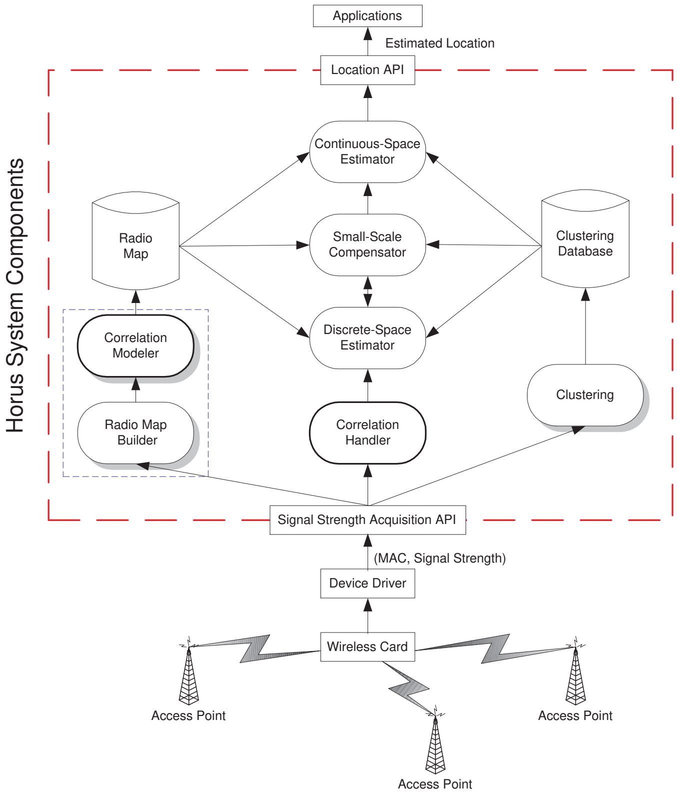

Fig. 1. *Horus* Components: the arrows show information flow in the system. Shaded block represent modules used during the offline phase. In this paper, we describe the correlation modeling and handling components of the *Horus* system (shown in thick lines).

uses a moving time-average of multiple consecutive location estimates to obtain a better location estimate.

Our technique uses multiple samples in the *signal-strength space* to obtain a better location estimate. Moreover our technique can be used in conjunction with the physical-location space averaging to enhance their accuracy as discussed in Section VII.

The proposed technique is unique in using multiple samples in the signal-strength space to enhance the accuracy of *probabilistic* location determination techniques *taking into account the high correlation degree between samples from the same access point.*

III. Background

A. Overview of the Horus System

In this section, we present a brief overview of the *Horus* system [1]–[3]. Our goal is to provide context for the technique presented in Section V. *Horus* 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. Figure 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, do and 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. There are two modes for operation of the *Horus* system: one uses non-parametric distributions and the other uses parametric distributions. In this paper, we will use the parametric distribution mode in which the signal strength distributions are modeled using *Gaussian* distributions.

The *Clustering* module is used to group radio map locations based on the access points covering them. Clustering is used to reduce the computational requirements of the system and, hence, conserve power.

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.

The *Small-Scale Compensator* module handles the small-scale variation characteristics of the wireless channel [2].

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.

In this paper, we describe the *Correlation Modeling* and the *Correlation Handling* modules of the *Horus* system.

B. Time Series Analysis

The technique described in this paper treats the samples received form an access point as a time series and use time series based-techniques to analyze the correlation between the samples. A *time series* [14] is a set of observations generated sequentially in time. If the set is discrete, the time series is said to be discrete, otherwise, it is a continuous time series. We refer to successive equi-spaced samples from a discrete time series s as $s_1, s_2, ...$. A statistical phenomenon that evolves in time according to probabilistic laws is called a *stochastic process*. The time series to be analyzed may be thought of as a particular realization of the system under study. A *stationary* stochastic process is based on the assumption that the process is in a particular state of statistical equilibrium. More formally, a discrete process is strictly stationary if the joint probability of any set of observations must be unaffected by shifting all the times of the observation forward or backward by any integer amount.

Alg. 1 x= Horus_GetLocation (s, $\mathbb{X}$, P_RM)

Input:

s : Measured signal strength vector from k access points ($s = (s_1, ..., s_k)$).

$\mathbb{X}$: 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 $x \in \mathbb{X}$.

Output:

The location $x \in \mathbb{X}$ that maximizes $P(x/s)$.

```
Max ← 0
for l ∈ 𝕏 do
    P ← ∏_{i=1}^{k} P_RM(s_i, i, l)
    if P > Max then
        x ← l
        Max ← p
    end if
end for
```

Autoregressive models are stochastic models used to analyze stochastic time series. In these models, the current values of the process is expressed as a finite, linear aggregate of previous values of the process and a noise v_t. Therefore, if we denote the values of the process as $s_t, s_{t-1}, s_{t-2}, ...$, then

$$s_t - \bar{s} = (\phi_1.s_{t-1} - \bar{s}) + (\phi_2.s_{t-2} - \bar{s}) + ... + (\phi_p.s_{t-p} - \bar{s}) + v_t \quad (1)$$

is called an autoregressive process of order p, where $\bar{s}$ is the average of the process.

In this paper, we treat the signal strength samples from an access point as a discrete stationary time series. We model this time series using a first order autoregressive model. To the best of our knowledge, this is the first work to apply time series techniques to the analysis of 802.11 signal strength characteristics.

IV. Signal Strength Temporal Characteristics

In this section, we present the temporal characteristics of the signal strength received from an access point and discuss how they affect the estimation of the user location. For a discussion of spatial characteristics, the reader is referred to [2].

A. Received Signal Strength Variations

Figure 2 shows the normalized histogram of the signal strength received from an access point at a fixed position. The figure shows that the measured signal strength at a fixed position varies over time and the variations can be as large as 10 dBm. This time variation of the channel can be due to changes in the physical environment such as people movement [15].

These variations suggests that depending on a single sample for estimating the user location may lead to inaccurate results if this sample comes from the tail of the distribution. This motivates the need for the techniques that are based on using more than one sample in estimating the user location.

B. Samples Correlation

Figure 3 shows the autocorrelation function of the samples collected from one access point (one sample per second) at a fixed position. The figure shows that the autocorrelation of consecutive samples ($lag = 1$) is as high as 0.9. This is a typical value for all the access points we experimented with. This high autocorrelation is expected as over a short period of time the signal strength received form an access point at a particular point is relatively stable (modulo the changes in the environment discussed in Section IV-A).

This high autocorrelation value should be considered when using the methods that use multiple samples suggested in the previous section, especially for *probabilistic* location determination techniques. Figure 6 shows the effect of averaging samples on the accuracy of a **probabilistic** WLAN location determination system that assumes the *independence* of samples[1]. The figure shows that although averaging increases accuracy, the wrong independence assumption leads to increasing average distance error *increases* as the number of averaged samples increases. The goal of this paper is to take the high samples correlation into account to further enhance the performance of probabilistic WLAN location determination systems.

[1]The figure is discussed in more details in Section VI.

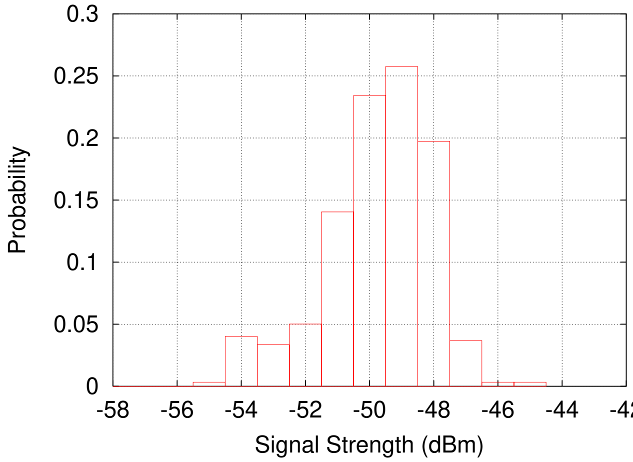

Fig. 2. An example of a normalized histogram of the signal strength of an access point.

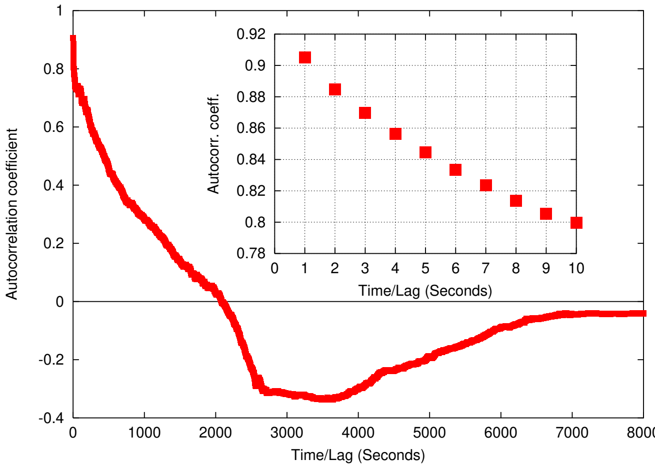

Fig. 3. An example of the autocorrelation between samples from an access point. The sub-figure shows the autocorrelation for the first 10 lags.

V. Handling Samples Correlation

This section describe an autoregressive model that capture the autocorrelation between samples from the same access point. Following that, we present a technique that uses this model to calculate the distribution of the average of n correlated samples. Finally, we modify the *Horus* location determination system to incorporate the new technique.

A. Autoregressive Model

Let s_t be the *stationary* time series representing the samples from an access point where t is the discrete time index. s_t can be represented as a *first order* autoregressive model as:

$$s_t = \alpha.s_{t-1} + (1-\alpha).v_t \quad ; 0 \leq \alpha \leq 1 \tag{2}$$

where v_t is a noise process, independent from s_t, and α is a parameter that determines the degree of autocorrelation of the original samples. Moreover, different samples from v_t are i.i.d.'s[2].

The model in Equation 2 states that the current signal strength value (s_t) is an linear aggregate of the previous signal strength value (s_{t-1}) and an independent noise value (v_t). The parameter α gives flexibility to the model as it can be used to determine the degree of autocorrelation of the original process. For example, if α is zero, the samples of the process s_t are i.i.d.'s, whereas if α is one the original samples are identical (autocorrelation=1). In the following subsections we describe some properties of the autoregressive model that we will use in the rest of the paper.

1) Relation Between the Mean of s_t and v_t: We can see from Equation 2 that $E(s_t) = E(v_t)$. The two processes have the same mean.

2) Relation Between the Variance of s_t and v_t: The relation between the variance of the original and noise processes can be obtained as follows:

$$\begin{aligned} Var(s_t) &= Var(\alpha.s_{t-1} + (1-\alpha).v_t) \\ &= \alpha^2.Var(s_{t-1}) + (1-\alpha)^2.Var(v_t) \\ &(s_t \text{ and } v_t \text{ independent}) \end{aligned} \tag{3}$$

Since the samples of s_t are identically distributed (stationary process), $Var(s_t) = Var(s_{t-1}) = Var(s)$. Therefore equation 3 can be rewritten as:

$$(1-\alpha^2)Var(s) = (1-\alpha)^2.Var(v_t) \tag{4}$$

therefore

$$\begin{aligned} \frac{Var(v_t)}{Var(s)} &= \frac{1-\alpha^2}{(1-\alpha)^2} \\ &= \frac{1+\alpha}{1-\alpha} \end{aligned} \tag{5}$$

[2]This model is equivalent to the one given in Equation 1.

3) Relation Between s_t and s_0: We start from equation 2.

$$\begin{aligned} s_t &= \alpha.s_{t-1} + (1-\alpha)v_t \\ &= \alpha^2.s_{t-2} + \alpha.(1-\alpha)v_{t-1} + (1-\alpha)v_t \\ &= \alpha^3.s_{t-3} + \alpha^2.(1-\alpha)v_{t-2} \\ &\quad + \alpha.(1-\alpha)v_{t-1} + (1-\alpha)v_t \\ &\vdots \\ &= \alpha^t.s_0 + (1-\alpha).\sum_{i=1}^{t} \alpha^{t-i}.v_i \end{aligned} \tag{6}$$

B. Estimating the Value of α

In this section, we show that α value can be approximated using the autocorrelation coefficient with lag one (r_1). r_1 is estimated from a sample of size N as [14]:

$$r_1 = \frac{\sum_{t=1}^{N-1} [s_t - \bar{s}].[s_{t+1} - \bar{s}]}{\sum_{t=1}^{N} [s_t - \bar{s}]^2} \tag{7}$$

where $\bar{s}$ is the expected value of process s.

For large values of α (close to one), Equation 2 can be approximated as:

$$s_t \approx \alpha.s_{t-1} \tag{8}$$

Substituting Equation 8 in Equation 7 yields:

$$\begin{aligned} r_1 &\approx \frac{\sum_{t=1}^{N-1} [s_t - \bar{s}].[\alpha.s_t - \bar{s}]}{\sum_{t=1}^{N} [s_t - \bar{s}]^2} \\ &\approx \frac{\sum_{t=1}^{N-1} [s_t - \bar{s}].[\alpha.(s_t - \bar{s}) - (1-\alpha).\bar{s}]}{\sum_{t=1}^{N} [s_t - \bar{s}]^2} \\ &\approx \frac{\alpha.\sum_{t=1}^{N-1} [s_t - \bar{s}]^2}{\sum_{t=1}^{N} [s_t - \bar{s}]^2} \quad (\alpha \text{ close to } 1) \end{aligned} \tag{9}$$

For large N, Equation 9 can be rewritten as:

$$r_1 \approx \alpha \tag{10}$$

Therefore for a large value of α and N, as is the case here, α can be estimated using the autocorrelation coefficient with lag one.

C. Distribution of the Average of n Correlated Samples

In this section, we obtain the mean and variance of the samples of a new process whose samples are the average of n samples from the original process.

1) Mean of the Distribution of the Average of n Samples: We use $A(n)$ to denote the random variable whose value is the average of n samples (from $t = 0$ to $t = n-1$) of the original process s_t, $n > 1$. Since

$$A(n) = \frac{1}{n}.\sum_{j=0}^{n-1} s_j \tag{11}$$

therefore, $E(A(n)) = E(s_t)$. The mean of the distribution of the average of n samples is the same as the mean of the distribution of each sample.

2) Variance of the Distribution of the Average of n Samples: From equation 6, $A(n)$ can be written as:

$$\begin{aligned} A(n) &= \frac{1}{n}.\sum_{j=0}^{n-1}\left\{\alpha^j.s_0 + (1-\alpha).\sum_{i=1}^{j}\alpha^{j-i}.v_i\right\} \\ &= \frac{1}{n}.\left\{\frac{1-\alpha^n}{1-\alpha}.s_0 + (1-\alpha).\sum_{j=1}^{n-1}\sum_{i=1}^{j}\alpha^{j-i}.v_i\right\} \end{aligned} \tag{12}$$

therefore,

$$\begin{aligned} Var\big(A(n)\big) &= \frac{1}{n^2}.\Big\{(\frac{1-\alpha^n}{1-\alpha})^2.Var(s_0) \\ &\quad + (1-\alpha)^2.\sum_{j=1}^{n-1}\sum_{i=1}^{j}\alpha^{2.(j-i)}.Var(v_i)\Big\} \end{aligned} \tag{13}$$

from equation 5

$$\begin{aligned} Var\big(A(n)\big) &= \frac{Var(s_0)}{n^2}.\Big\{(\frac{1-\alpha^n}{1-\alpha})^2 \\ &\quad + (1-\alpha^2).\sum_{j=1}^{n-1}\sum_{i=1}^{j}\alpha^{2.(j-i)}\Big\} \\ &= \frac{Var(s_0)}{n^2}.\Big\{(\frac{1-\alpha^n}{1-\alpha})^2 \\ &\quad + (1-\alpha^2).\sum_{j=1}^{n-1}\frac{1-\alpha^{2.j}}{1-\alpha^2}\Big\} \\ &= \frac{Var(s_0)}{n^2}.\Big\{(\frac{1-\alpha^n}{1-\alpha})^2 + n - 1 \\ &\quad - \alpha^2.\frac{1-\alpha^{2.(n-1)}}{1-\alpha^2}\Big\} \end{aligned} \tag{14}$$

Since s_t is a stationary process, $Var(s_0) = Var(s)$ and the final relation between $Var\big(A(n)\big)$ and $Var(s)$ is:

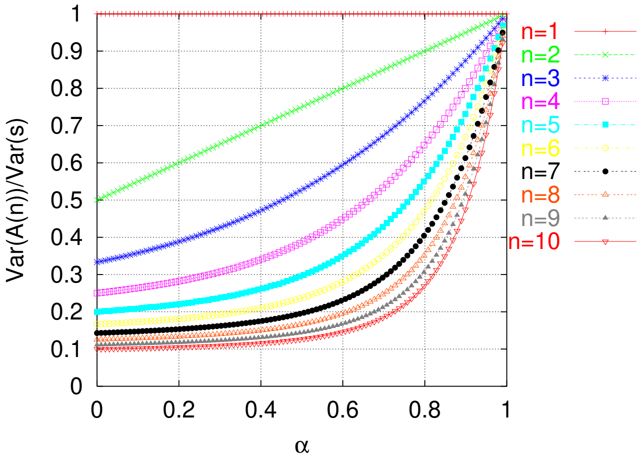

Fig. 4. Ratio between the variances of the averaging process and the original process for different values of α and n

$$\begin{aligned} Var\big(A(n)\big) = \frac{Var(s)}{n^2}.\Big\{(\frac{1-\alpha^n}{1-\alpha})^2 + n - 1 \\ - \alpha^2.\frac{1-\alpha^{2.(n-1)}}{1-\alpha^2}\Big\} \end{aligned} \tag{15}$$

Note that when $\alpha = 0$ (i.e. the samples of s_t are independent), Equation 16 reduces to:

$$Var\big(A(n)\big) = \frac{Var(s)}{n} \tag{16}$$

Figure 4 shows the ratio between the variance of the averaging process and original process for different values of α and n. The variance of the averaging process,$Var(A(n))$, is always less than or equal to the variance of the original process,$Var(s)$, being equal in case α equals one. Intuitively, the lower the variance of the signal strength distribution at each location, the better the ability to discriminate between different locations and the better the accuracy.

D. Modified Horus Algorithm

In this section, we use the results of the previous section to obtain the distribution of the average of n correlated samples. We use this value to determine the most probable user location. We assume that the *Horus* system is running in the parametric mode where the signal strength distribution follows a *Gaussian* distribution [3], [11], [12]. Since the individual distribution of each sample follows a *Gaussian* distribution, the probability distribution of the average of n samples follows a *Gaussian* distribution whose mean and variance can be obtained using the results in Section V-C.

The technique works as follows:

- *Offline phase*: the system calculates the parameters of the distribution of the average of n samples for each access point in the radio map.
- *Online phase*: Given n samples from an access point, the algorithm obtains their average and calculate the probability of each radio map location given this value of the average using the distribution of the average of n samples calculated during the offline phase.

Algorithm 2 shows the details of the modified *Horus* algorithm. Note that the value of α is implicitly used in the online phase as the distribution of the average of n samples depends on the value of α as discussed in Section V-C.

Alg. 2 x= Corr_Horus_GetLocation (n, S, $\mathbb{X}$, P_RM)

Input:

n : Number of samples from each access point.
S : Measured signal strength vectors from k access points ($S = (\bar{s_1}, ..., \bar{s_k})$). Each $\bar{s_i}, 1 \leq i \leq k$ is a vector containing n samples from access point i.
$\mathbb{X}$: Radio map locations.
RM : A radio map based function, where $P\_RM(s_a, a, x)$ returns the probability of the average, of the n samples received from access point a at location $x \in \mathbb{X}$, being s_a.

Output:

The location $x \in \mathbb{X}$ that maximizes $P(x/S)$.

```
for i = 1..k do
    Avg(i) ← average(s̄_i)
end for
Max ← 0
for l ∈ X do
    P ← ∏_{i=1}^{k} P_RM(Avg(i), i, l)
    if P > Max then
        x ← l
        Max ← p
    end if
end for
```

VI. Experimental Evaluation

In this section we present the result of implementing the correlation handling technique in the context of the *Horus* system.

A. Experimental Testbed

We performed our experiment in the south wing of the fourth floor of the Computer Science Department building. The layout of the floor is shown in Figure 5. The wing has a dimension of 224 feet by 85.1 feet. The technique was tested in the Computer Science Department wireless network. The entire wing is covered by 12 access points installed in the third and fourth floors of the building.

For building the radio map, we took the radio map locations on the corridors on a grid with cells placed 5 feet apart (the corridor's width is 5 feet). We have a total of 110 locations along the corridors. On the average, each location is covered by 4 access points. The value of α, autocorrelation degree, for these access points was approximately 0.9 for all access points.

Using the device driver and the API we developed [16], we collected 300 samples at each location, one sample per second. The cards used were Lucent Orinoco silver NICs supporting up to 11 Mbit/s data rate [17]. *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.*

B. Results

We start by showing the effect of the wrong independence assumption on the performance of the original *Horus* system. Figure 6 shows the average distance error for different values of n. We can see that averaging can significantly improves performance (average error decreases by about 2 feet from $n = 1$ to $n = 2$). However, as the number of averaged samples increases, the performance degrades. The minimum value at $n = 2$ can be explained by noting that there are two opposing factors affecting the system accuracy:

1) as the number of averaged samples n increases, the accuracy of the system should increase.
2) as n increases, the estimation of the distribution of the average of the n samples becomes worse due to the wrong independence assumption.

At low values of n ($n = 1, 2$) the first factor is the dominating factor and hence the accuracy increases. Starting from $n = 3$, the effect of the bad estimation of the distribution becomes the dominating factor and accuracy degrades.

Figure 7 shows the average distance error for different values of α and n. The figure shows that as the value of α, used in calculating the parameters of the distribution

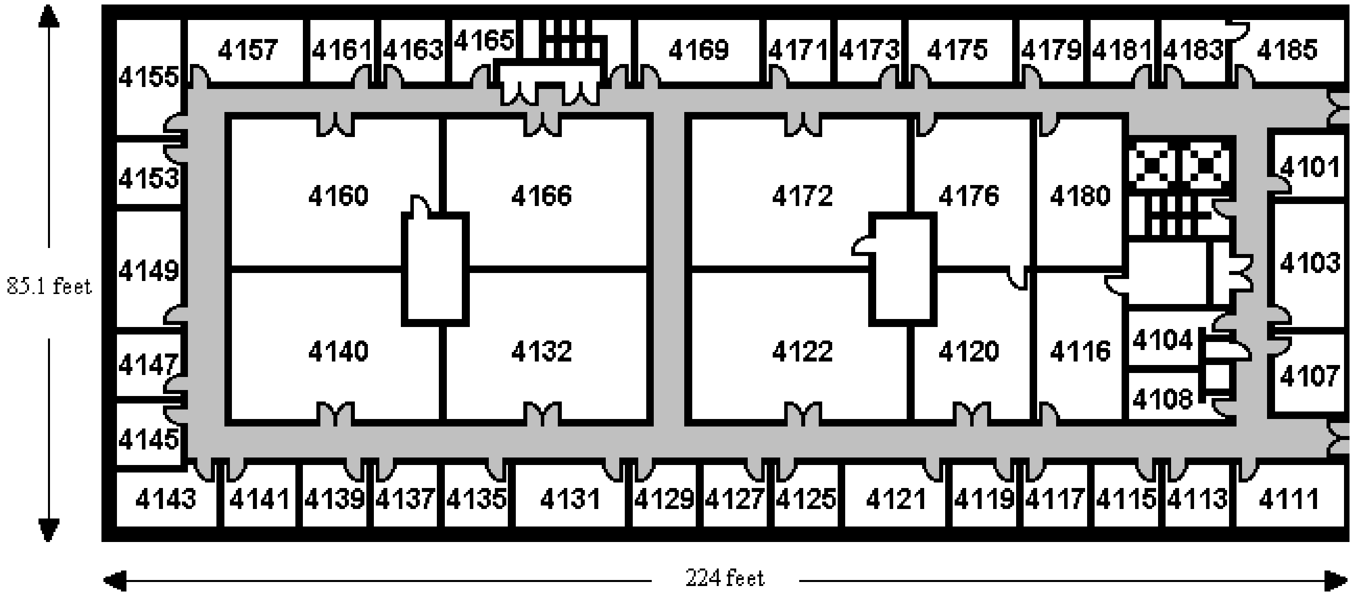

Fig. 5. Plan of the south wing of the 4th floor of the Computer Science Department building where the experiment was conducted. Readings were collected in the corridors (shown in gray).

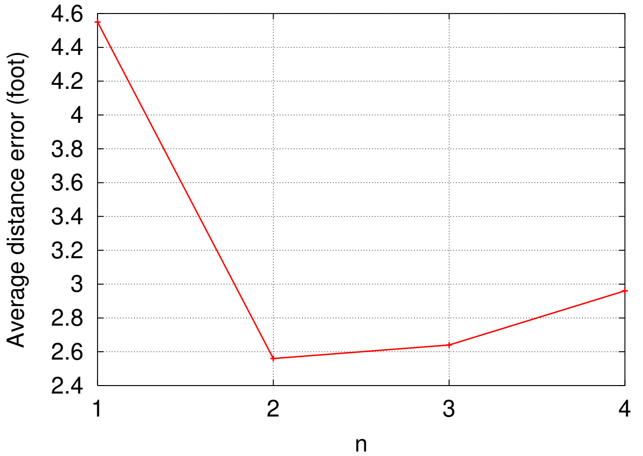

Fig. 6. Effect of the wrong independence assumption on the average distance error. As the number of averaged samples increases, the average system error increases.

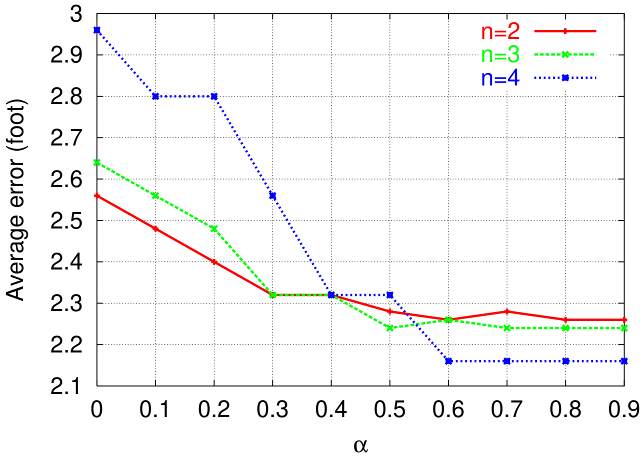

Fig. 7. Average distance error for different values of α and n. As the value of α approaches the true value of 0.9, the system performance increases. The case for $n = 1$ (original *Horus* system performance) is shown in Figure 6 for clarity.

of the average of n samples, approaches the true α value (0.9), the system accuracy increases.

Note that at low values of α averaging more samples lead to worse accuracy, as shown in Figure 6, till we reach a switch-over point at about $\alpha = 0.4$ where averaging more samples starts to give better accuracy. Using the modified technique, the system can achieve an average accuracy of about 2.15 feet, better than the original system by more than 2.4 feet.

VII. Discussion and Conclusions

The main contribution of this paper is three fold: (a) We applied the time series analysis techniques to modeling the behavior of signal strength samples from an access point, (b) we presented a technique that uses multiple samples from the same access point, taking high correlation into account, to enhance the accuracy of probabilistic WLAN location determination systems, and (c) we analyzed the performance of the proposed technique by implementing it in the context of the *Horus* system.

We showed that the samples autocorrelation can be as large as 0.9 and therefore it becomes crucial to take this high autocorrelation into account when designing location determination algorithms that uses more than one samples. We described an autoregressive model to capture the autocorrelation between samples from the same access point, showed how to estimate its parameters, and derived some useful properties of the model.

Based on the autoregressive model, we presented a technique that uses the average of n samples from the same access point, taking samples autocorrelation into account, to enhance the accuracy of the *Horus* system. The results show that the average distance accuracy is enhanced by more than 2.4 feet (50%) using the modified technique. Assuming independence of samples leads to degraded accuracy as the number of averaged samples increases as the estimate distribution of the average of n samples becomes worse with increasing n.

For the proposed technique, as n increases the accuracy of the system is enhanced. However, a side effect of this increased accuracy is that the latency of calculating the location estimate increases with the increases of the number of samples required. In general, we have a trade-off between the accuracy required and the latency of location estimate. The higher the required accuracy, the higher n and the higher the latency to obtain the location estimate. This decision is dependent on the application in use.

Latency can be reduced by presenting the location estimate incrementally using one sample at a time. The system need not to wait till it acquires the n samples all at once. Instead, it can give a more accurate estimate of the location as more samples become available by reporting the estimated location given the partial samples it has. In this mode, the system can be incorporated with a physical-location space moving averaging method to further enhance the accuracy.

Other factors that affect the choice of the value of n are the user mobility rate and the sampling rate. The higher the user mobility rate or the sampling rate, the lower the value of n.

The computational overhead of the modified technique is minimal. The system calculates the parameters of the distribution of the average of n samples in the offline phase, which is done only once. The only additional requirement in the online phase is calculating the average which involves n addition operations and one division. This is amortized over the number of locations in the radio map.

We believe that the model and the technique presented in the paper are general and can be applied to other probabilistic WLAN location determination techniques to enhance their accuracy.

REFERENCES

[1] M. Youssef, A. Agrawala, and A. U. Shankar, "WLAN Location Determination via Clustering and Probability Distributions," in *IEEE PerCom 2003*, March 2003.

[2] M. Youssef and A. Agrawala, "Small-Scale Compensation for WLAN Location Determination Systems," in *IEEE WCNC 2003*, March 2003.

[3] M. Youssef, A. Agrawala, A. U. Shankar, and Sam H. Noh, "A Probabilistic Clustering-Based Indoor Location Determination System," Tech. Rep. UMIACS-TR 2002-30 and CS-TR 4350, University of Maryland, College Park, March 2002, http://www.cs.umd.edu/Library/TRs/.

[4] The Institute of Electrical and Electronics Engineers, Inc., "IEEE Standard 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," 1999.

[5] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Tech. Rep. Dartmouth Computer Science Technical Report TR2000-381, 2000.

[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] P. Bahl, V. N. Padmanabhan, and A. Balachandran, "Enhancements to the RADAR User Location and Tracking System," Tech. Rep. MSR-TR-00-12, Microsoft Research, February 2000.

[8] 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.

[9] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Location Service for Wireless Network Environments," *Ubiquitous Computing 2001*, September 2001.

[10] P. Castro and R. Muntz, "Managing Context for Smart Spaces," *IEEE Personal Communications*, OCTOBER 2000.

[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*, vol. 9, no. 3, July 2002.

[12] T. Roos, P. Myllymaki, and H. Tirri, "A Statistical Modeling Approach to Location Estimation," *IEEE Transactions on Mobile Computing*, vol. 1, no. 1, pp. 59–69, January-March 2002.

[13] 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.

[14] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forcasting and Control*, Prentice Hall, third edition, 1994.

[15] T. S. Rappaport, *Wireless Communications Principles and Practice*, Pearson Education, second edition, 2002.

[16] "http://www.cs.umd.edu/users/moustafa/Downloads.html," .

[17] "http://www.orinocowireless.com," .