On the Optimality of WLAN Location Determination Systems

Moustafa Youssef
Department of Computer Science
University of Maryland
College Park, Maryland 20742
Email: moustafa@cs.umd.edu

Ashok Agrawala
Department of Computer Science
University of Maryland
College Park, Maryland 20742
Email: agrawala@cs.umd.edu

***Abstract*— This paper presents a general analysis for the performance of WLAN location determination systems. In particular, we present an analytical method for calculating the average distance error and probability of error of WLAN location determination systems. These expressions are obtained with no assumptions regarding the distribution of signal strength or the probability of the user being at a specific location, which is usually taken to be a uniform distribution over all the possible locations in current WLAN location determination systems. We use these expressions to find the optimal strategy to estimate the user location and to prove formally that probabilistic techniques give more accuracy than deterministic techniques, which has been taken for granted without proof for a long time. The analytical results are validated through simulation experiments and we present the results of testing actual WLAN location determination systems in an experimental testbed.**

Keywords – *Analytical analysis, optimal WLAN positioning strategy, simulation experiments, WLAN location determination*

I. Introduction

WLAN location determination systems use the popular 802.11 [10] network infrastructure to determine the user location without using any extra hardware. This makes these systems attractive in indoor environments where traditional techniques, such as the Global Positioning System (GPS) [5], fail to work or require specialized hardware. Many applications have been built on top of location determination systems to support pervasive computing. This includes [4] location-sensitive content delivery, direction finding, asset tracking, and emergency notification.

In order to estimate the user location, a system needs to measure a quantity that is a function of distance. Moreover, the system needs one or more reference points to measure the distance from. In case of the GPS system, the reference points are the satellites and the measured quantity is the time of arrival of the satellite signal to the GPS receiver, which is directly proportional to the distance between the satellite and the GPS receiver. In case of WLAN location determination systems, the reference points are the access points and the measured quantity is the signal strength, which decays logarithmically with distance in free space. Unfortunately, in indoor environments, the wireless channel is very noisy and the radio frequency (RF) signal can suffer from reflection, diffraction, and multipath effect [9], [12], which makes the signal strength a complex function of distance. To overcome this problem, WLAN location determination systems tabulate this function by sampling it at selected locations in the area of interest. This tabulation has been know in literature as the radio map, which captures the signature of each access point at certain points in the area of interest.

WLAN location determination systems usually work in two phases: offline phase and location determination phase. During the offline phase, the system constructs the radio-map. In the location determination phase, the vector of samples received from each access point (each entry is a sample from one access point) is compared to the radio-map and the "nearest" match is returned as the estimated user location. Different WLAN location determination techniques differ in the way they construct the radio map and in the algorithm they use to compare a received signal strength vector to the stored radio map in the location determination phase.

In this paper, we present a *general* analysis of the performance of WLAN location determination systems. In particular, we present a general analytical expression for the average distance error and probability of error of WLAN location determination systems. These expression are obtained with *no assumptions regarding the distribution of signal strength or user movement profile*. We use these expressions to find the *optimal* strategy to use during the location determination phase to estimate the user location. These expressions also help to prove ***formally*** that probabilistic techniques give more accuracy than deterministic techniques, which has been taken for granted without proof for a long time. We validate our analysis through simulation experiments and discuss how well it models actual environments. For the rest of the paper we will refer to the probability distribution of the user location as the *user profile*.

To the best of our knowledge, our work is the first to analyze the performance of WLAN location systems analytically and provide the optimal strategy to select the user location.

The rest of this paper is structured as follows. section II summarizes the previous work in the area of WLAN location determination systems. section III presents the analytical analysis for the performance of the WLAN location determination systems. In section IV, we validate our analytical analysis through simulation and measurement experiments. Section V concludes the paper and presents some ideas for future work.

II. Related Work

Radio map-based techniques can be categorized into two broad categories: deterministic techniques and probabilistic techniques. *Deterministic techniques*, such as [2], [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 [2] the authors use nearest neighborhood techniques to infer the user location. On the other hand, *probabilistic techniques*, such as [3], [6], [7], [13], [14], 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 *Horus* system from the University of Maryland [14], [15] uses the stored radio map to find the location that has the maximum probability given the received signal strength vector.

All these systems base their performance evaluation on experimental testbeds which may not give a good idea on the performance of the algorithm in different environments. The authors in [7], [14], [15] showed that their probabilistic technique outperformed the deterministic technique of the *Radar* system [2] in a *specific* testbed and conjectured that probabilistic techniques should outperform deterministic techniques. This paper presents a general *analytical* method for analyzing the performance of different techniques. We use this analysis method to provide a formal proof that probabilistic techniques outperform deterministic techniques. Moreover, we show the optimal strategy for selecting locations in the location determination phase.

III. Analytical Analysis

In this section, we give an analytical method to analyze the performance of WLAN location determination techniques. We start by describing the notations used throughout the paper. We provide two expressions: one for calculating the average distance error of a given technique and the other for calculating the probability of error (i.e. the probability that the location technique will give an incorrect estimate).

A. Notations

We consider an area of interest whose radio map contains N locations. We denote the set of locations as $\mathbb{L}$. At each location, we can get the signal strength from k access points. We denote the k-dimensional signal strength space as $\mathbb{S}$. Each element in this space is a k-dimensional vector whose entries represent the signal strength reading from different access points. Since the signal strength returned from the wireless cards are typically integer values, the signal strength space $\mathbb{S}$ is a discrete space. For a vector $s \in \mathbb{S}$, $f^*_{\mathcal{A}}(s)$ represents the estimated location returned by the WLAN location determination technique $\mathcal{A}$ when supplied with the input s. For example, in the *Horus* system [14], [15], $f^*_{\text{Horus}}(s)$ will return the location $l \in \mathbb{L}$ that maximizes $P(l/s)$. Finally, we use *Euclidean*(l_1, l_2) to denote the Euclidean distance between two locations l_1 and l_2.

B. Average Distance Error

We want to find the average distance error (denoted by $E(\text{DErr})$). Using conditional probability, this can be written as:

$$\begin{aligned} E(\text{DErr}) \quad = \quad & \sum_{l \in \mathbb{L}} E(\text{DErr}/l \text{ is the correct user location}) \\ & .P(l \text{ is the correct user location}) \qquad (1) \end{aligned}$$

where $P(l \text{ is the correct user location})$ depends on the user profile.

We now proceed to calculate $E(\text{DErr}/l \text{ is the correct user location})$. Using conditional probability again:

$$
\begin{aligned}
&E(\text{DErr}/l \text{ is the correct user location}) \\
&= \sum_{s\in\mathbb{S}} E(\text{DErr}/s, l \text{ is the correct user location}) \\
&.P(s/l \text{ is the correct user location}) \\
&= \sum_{s\in\mathbb{S}} \text{Euclidean}(f^*_{\mathcal{A}}(s), l) \\
&.P(s/l \text{ is the correct user location})
\end{aligned} \tag{2}
$$

where Euclidean$(f^*_{\mathcal{A}}(s), l)$ represents the Euclidean distance between the estimated location and the correct location.

Equation 2 says that to get the expected distance error given we are at location l, we need to get the weighted sum, over all the possible signal strength values $s \in \mathbb{S}$, of the Euclidean distance between the estimated user location ($f^*_{\mathcal{A}}(s)$) and the actual location l.

Substituting equation 2 in equation 1 we get:

$$
\begin{aligned}
E(\text{DErr}) \;=\; & \sum_{s\in\mathbb{S}} \sum_{l\in\mathbb{L}} \text{Euclidean}(f^*_{\mathcal{A}}(s), l) \\
& .P(s/l \text{ is the correct user location}) \\
& .P(l \text{ is the correct user location})
\end{aligned} \tag{3}
$$

Note that the effect of the location determination technique is summarized in the function $f^*_{\mathcal{A}}$. We seek to find the function that minimizes the probability of error. We differ the optimality analysis till we present the *probability of error* analysis.

C. Probability of Error

In this section, we want to find an expression for the probability of error which is the probability that the location determination technique will return an incorrect estimate. This can be obtained from equation 3 by noting that every non-zero distance error (represented by the function Euclidean$(f^*_{\mathcal{A}}(s), l)$) is considered an error. More formally, we define the function:

$$
g(x) = \begin{cases} 0 & : \quad x = 0 \\ 1 & : \quad x > 0 \end{cases}
$$

The probability of error can be calculated from equation 3 as:

$$
\begin{aligned}
P(\text{Error}) \;=\; & \sum_{s\in\mathbb{S}} \sum_{l\in\mathbb{L}} \text{g(Euclidean}(f^*_{\mathcal{A}}(s), l)) \\
& .P(s/l \text{ is the correct user location}) \\
& .P(l \text{ is the correct user location})
\end{aligned} \tag{4}
$$

In the next section, we will present a property of the term g(Euclidean$(f^*_{\mathcal{A}}(s), l)$) and use this property to get the optimal strategy for selecting the location.

D. Optimality

We will base our optimality analysis on the probability of error.

Lemma 1: For a given signal strength vector s, g(Euclidean$(f^*_{\mathcal{A}}(s), l)$) will be zero for only one location $l \in \mathbb{L}$ and one for the remaining $N-1$ locations.

Proof: The proof can be found in [11] and have been removed for space constraints. ■

The lemma states that only one location will give a value of zero for the function g(Euclidean$(f^*_{\mathcal{A}}(s), l)$) in the inner sum. This means that the optimal strategy should select this location in order to minimize the probability of error. This leads us to the following theorem.

Theorem 1 (Optimal Strategy): Selecting the location l that maximizes the probability $P(s/l).P(l)$ is both a necessary and sufficient condition to minimize the probability of error.

Proof: The proof can be found in [11]. ■

Theorem 1 suggests that the optimal location determination technique should store in the radio map the signal strength distributions to be able to calculate $P(s/l)$. Moreover, the optimal technique needs to know the user profile in order to calculate $P(l)$.

Corollary 1: Deterministic techniques are not optimal.

Proof: The proof can be found in [11]. ■

Note that we did not make any assumption about the independence of access points, user profile, or signal strength distribution in order to get the optimal strategy.

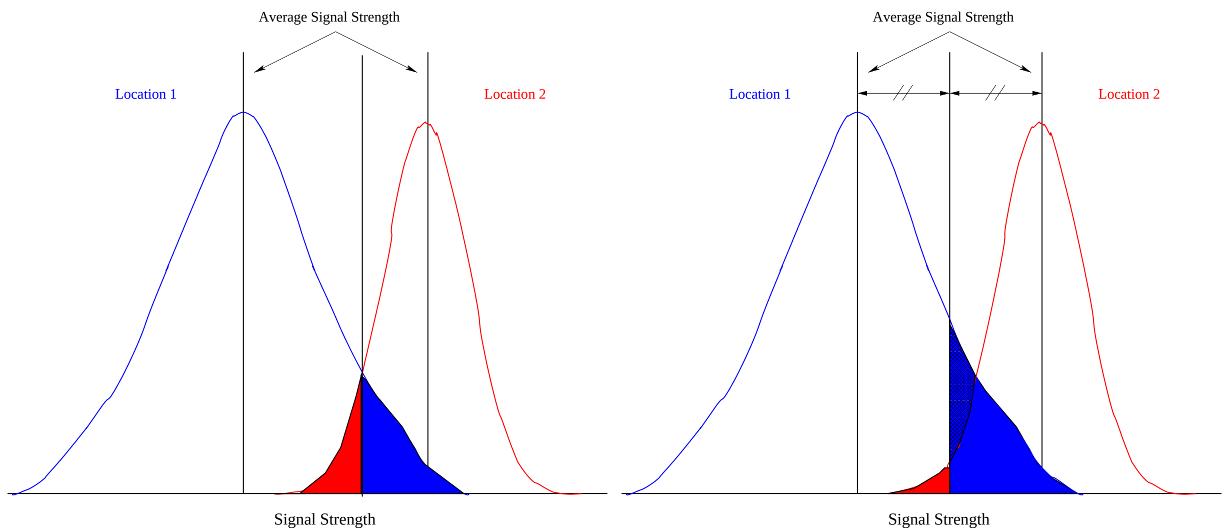
A major assumption by most of the current WLAN location determination systems is that all user locations are equi-probable. In this case, $P(l) = \frac{1}{N}$ and Theorem 1 can be rewritten as:

Theorem 2: If the user is equally probable to be at any location of the radio map locations $\mathbb{L}$, then selecting the location l that maximizes the probability $P(s/l)$ is both a necessary and sufficient condition to minimize the probability of error.

Proof: The proof is a special case of the proof of Theorem 1. □ ■

This means that, for this special case, it is sufficient for the optimal technique to store the histogram of signal strength at each location. This is exactly the technique used in the *Horus* system [14], [15].

Figure 1 shows a simplified example illustrating the intuition behind the analytical expressions and the theorems. In the example, we assume that there are only two locations in the radio map and that at each location only one access point can be heard whose signal strength, for



(a) Horus System (b) Radar System

Fig. 1. Expected error for the special case of two locations

simplicity of illustration, follows a continuous distribution. The user can be at any one of the two locations with equal probability. For the *Horus* system (Figure 1.a), consider the line that passes by the point of intersection of the two curves. Since for a given signal strength the technique selects the location that has the maximum probability, the error if the user is at location 1 is the area of curve 1 to the right of this line. If the user is at location 2, the error is the area of curve 2 to the left of this line. The expected error probability is half the sum of these two areas as the two locations are equi-probable. This is the same as half the area under the minimum of the two curves (shaded in figure).

For the *Radar* system (Figure 1.b), consider the line that bisects the signal strength space between the two distribution averages. Since for a given signal strength the technique selects the location whose average signal strength is closer to the signal strength value, the error if the user is at location 1 is the area under curve 1 to the right of this line. If the user is at location 2, the error is the area under curve 2 to the left of this line. The expected error probability is half the sum of these two areas as the two locations are equi-probable (half the shaded area in the figure).

From Figure 1, we can see that the *Horus* system outperforms the *Radar* system since the expected error for the former is less than the later (by the hashed area in Figure 1.b). The two systems would have the same expected error if the line bisecting the signal strength space of the two averages passes by the intersection point of the two curves. This is not true in general. This has been proved formally in the above theorems.

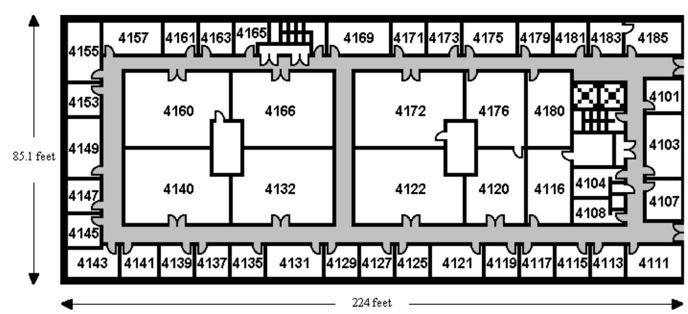


Fig. 2. Plan of the floor where the experiment was conducted. Readings were collected in the corridors (shown in gray).

We provide simulation and experimental results to validate our results in section IV.

IV. Experiments

A. Testbed

We performed our experiment in a floor covering an 20,000 feet area. The layout of the floor is shown in Figure 2. Both techniques were 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.

We used the *mwvlan* driver and the *MAPI* API [1] to collect the samples from the access points.

B. Simulation Experiments

In this section, we validate our analytical results through simulation experiments. For this purpose, we chose to implement the *Radar* system [2] from Microsoft as a deterministic technique and the *Horus* system [14], [15] from the University of Maryland as a probabilistic technique that satisfy the optimality criteria as described in Theorem 2. We start by describing the experimental testbed that we use to validate our analytical results and evaluate the systems.

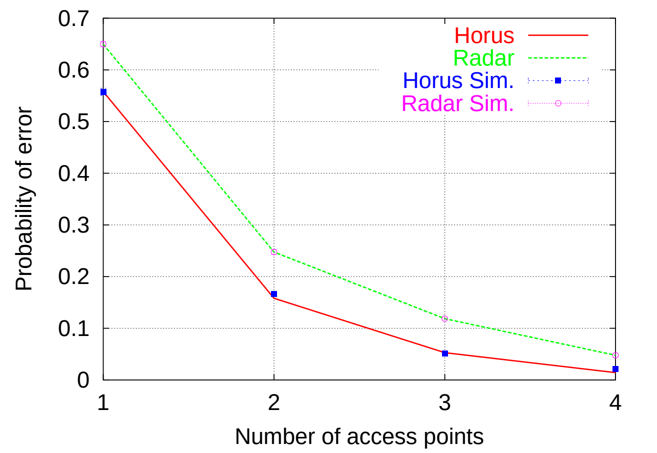
1) Simulator: We built a simulator that takes as an input the following parameters:

- the radio map locations coordinates.
- the signal strength distributions at each location from each access point.
- the distribution over the radio map locations that represents the steady state probability of the user being at each location (*user profile*).

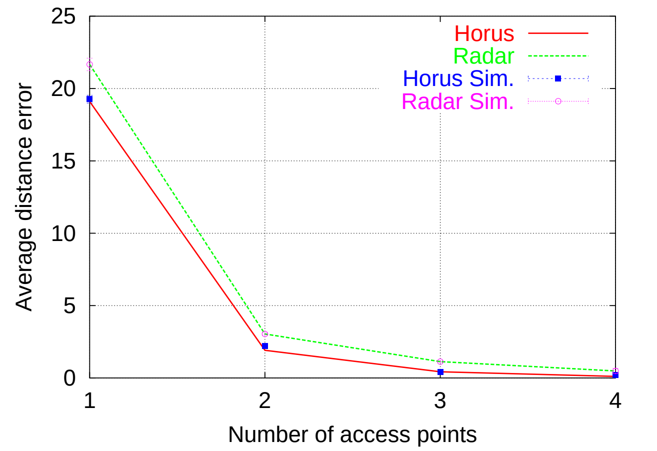
The simulator then chooses a location based on the user location distribution and generates a signal strength vector according to the signal strength distributions at this location. The simulator feeds the generated signal strength vector to the location determination technique. The estimated location is compared to the generated location to determine the distance error.

The next section analyze the effect of the uniform user profile on the performance of the location determination systems and validate our analytical results. The results for the heterogeneous profiles can be found in [11].

2) Uniform user location distribution: This is similar to the assumption taken by the *Horus* system. Therefore, the *Horus* system should give optimal results. Figures 3 shows the probability of error and average distance error (analytical and simulation results) respectively for the *Radar* and the *Horus* systems. The error bars represent the 95% confidence interval for the simulation experiments. The figure shows that the analytical expressions obtained are consistent with the simulation results. Moreover, the *Horus* system performance is better than the *Radar* system as predicted by Theorem 2. The *Horus* system performance is optimal under the uniform distribution of user location.



(a) Probability of error



(b) Average distance error

Fig. 3. Performance of the *Horus* and *Radar* systems under a uniform user profile (profile 1).

C. Measurements Experiments

In our simulations, we assumed that the test data follows the signal strength distributions exactly. This can be considered as the ideal case since in a real environment, the received signal may differ slightly form the stored signal strength distributions. Our results however are still valid and can be considered as an upper bound on the performance of the simulated systems. In order to confirm that, we tested the *Horus* system and the *Radar* system in an environment where the test set was collected on different days, time of day and by different persons than those in the training set.

Figure 4 shows the CDF of the distance error for the two systems. The figure shows that the *Horus* system (a probabilistic technique) significantly outperforms the *Radar* system (a deterministic technique) which confirms to our results.

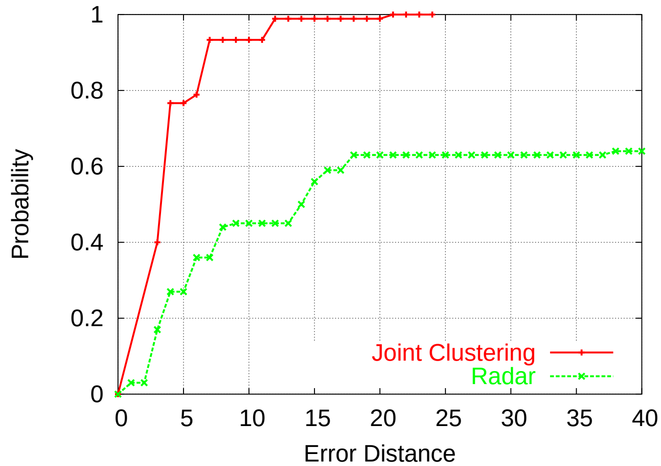


Fig. 4. CDF for the distance error for the two systems.

V. Conclusions and Future Work

We presented an analysis method for studying the performance of WLAN location determination systems. The method can be applied to any of the WLAN location determination techniques and does not make any assumptions about the signal strength distributions at each location, independence of access points, nor the user profile. Second, we studied the effect of the user profile on the performance of the WLAN location determination systems.

We used the analytical method to obtain the optimal strategy for selecting the user location. The optimal strategy must take into account the signal strength distributions at each location and the user profile. We validated the analytical results through simulation experiments.

In our simulations, we assumed that the test data follows the signal strength distributions exactly. This can be considered as the ideal case since in a real environment, the received signal may differ slightly form the stored signal strength distributions. Our results however are still valid and can be considered as an upper bound on the performance of the simulated systems. We confirmed that through actual implementation in typical environments.

For future work, the method can be extended to include other factors that affects the location determination process such as averaging multiple signal strength vectors to obtain better accuracy, using the user history profile, usually taken as the time average of the latest location estimates, and the correlation between samples from the same access points. cation determination systems.

Acknowledgment

This work was supported in part by the Maryland Information and Network Dynamics (MIND) Laboratory, its Founding Partner Fujitsu Laboratories of America, and by the Department of Defense through a University of Maryland Institute for Advanced Computer Studies (UMIACS) contract.

References

[1] http://www.cs.umd.edu/users/moustafa/Downloads.html.

[2] P. Bahl and V. N. Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System. In *IEEE Infocom 2000*, volume 2, pages 775–784, March 2000.

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

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

[5] P. Enge and P. Misra. Special issue on GPS: The Global Positioning System. *Proceedings of the IEEE*, pages 3–172, January 1999.

[6] 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.

[7] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen. A Probabilistic Approach to WLAN User Location Estimation. *International Journal of Wireless Information Networks*, 9(3), July 2002.

[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] W. Stallings. *Wireless Communications and Networks*. Prentice Hall, first edition, 2002.

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

[11] M. Youssef and A. Agrawala. On the Optimality of WLAN Location Determination Systems. Technical Report UMIACS-TR 2003-29 and CS-TR 4459, University of Maryland, College Park, March 2003.

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

[13] M. Youssef and A. Agrawala. Handling Samples Correlation in the Horus System. In *IEEE Infocom 2004*, March 2004.

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

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