A Locally Optimal Handoff Algorithm for Cellular Communications

Venugopal V. Veeravalli, *Member, IEEE*, and Owen E. Kelly, *Member, IEEE*

***Abstract*—The design of handoff algorithms for cellular communication systems based on mobile signal-strength measurements is considered. The design problem is posed as an optimization to obtain the best tradeoff between the *expected number of service failures* and *expected number of handoffs*, where a service failure is defined to be the event that the signal strength falls below a level required for satisfactory service to the subscriber. Based on dynamic programming arguments, an optimal solution is obtained, which, though impractical, can be used as a benchmark in the comparison of suboptimal schemes. A simple *locally optimal* handoff algorithm is derived from the optimal solution. Like the standard hysteresis algorithm, the locally optimal algorithm is characterized by a single threshold. A systematic method for the comparison of various handoff algorithms that are akin to the receiver operating characteristic (ROC) curves of radar detection is presented. Simulation results show that the locally optimal algorithm outperforms the hysteresis algorithm, especially in situations where accurate prediction of signal strength is possible. A straightforward technique for adapting the locally optimal algorithm to changing environments is suggested. That natural adaptability is the algorithm's principle advantage over current approaches.**

***Index Terms*—Handoff, handover, land mobile radio cellular systems.**

I. Introduction

A HANDOFF in cellular communication is the process whereby a *mobile subscriber* communicating with one *base station* is switched to another base station during a call. The design of reliable handoff algorithms is crucial to the operation of a cellular communication system and is especially important in microcellular systems, where the mobile may traverse several cells during a call.

Handoff decisions can be based on measurements[1] such as the signal strength, bit error rate, and estimated distance from base stations. In many systems, especially microcellular systems, signal strength may be the only reliable measurement that can be used [2]. Some measurements such as bit error rate are (either explicitly or implicitly) functions of the signal strength. If location information about the mobile is available independent from signal-strength measurements, then it can be used to improve handoff decisions derived from signal-strength measurements.

We focus on handoff algorithms based on signal-strength measurements. Consider the problem of designing an optimal handoff algorithm for a mobile moving between neighboring cells. Many criteria for determining the efficacy of a handoff algorithm are discussed in the literature [3]–[7] and may be used for optimal design. These include:

1) a measure of call quality, such as the received signal strength from the operative base station;
2) the total number of handoffs on a trajectory between neighboring cells;
3) the number of *unnecessary handoffs*, i.e., those made in situations, where the current base station would have continued to give satisfactory performance;
4) the number of *bad handoffs*, i.e., those made to a base station whose signal strength is below the satisfactory performance level;
5) the delay in making a handoff, which is sometimes defined to be the distance of the *crossover point*[2] from the cell boundary.

Considering these criteria, we can see that limiting the number of handoffs between cells [item 2)], while keeping the call quality high [item 1)] will generally eliminate those handoffs that do not enhance call quality. In a system that balances items 1) and 2), separate consideration of unnecessary and bad handoffs [items 3) and 4)] is not required. Concerning delay [item 5)], it can be argued that delay in handoff is undesirable only insofar as it impacts call quality. However, delay in handoff can create uplink interference to other mobiles in the system, which results in a network cost we do not consider in our analysis.[1] We limit our attention, therefore, to the first two criteria: call quality and total number of handoffs, but, clearly, these are conflicting criteria—improved call quality can usually be obtained only at the expense of an increase in the number of handoffs. Thus, a tradeoff must be made.

The standard approach to effecting a tradeoff between call quality and (expected) the number of handoffs has been through an *ad hoc* algorithm based on signal-strength *hysteresis*. The implicit measure of call quality used is simply the (average) value of the received signal strength from the operative base station. The hysteresis algorithm is designed so that handoff is made when the (averaged) signal strength

Manuscript received May 11, 1995; revised March 11, 1996. This work was presented in part at the 1995 International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC).

V. V. Veeravalli was with Rice University, Houston, TX 77005 USA. He is now with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: venu@ee.cornell.edu).

O. E. Kelly was with Rice University, Houston, TX 77005 USA. He is now with the Wireless Information Network laboratory, Rutgers University, Piscataway, NJ, 08855 USA (e-mail: oekelly@winlab.rutgers.edu).

Publisher Item Identifier S 0018-9545(97)03129-0.

[1] We restrict our attention to criteria based on degradation of the current channel. Handoff decisions can also be based on network criteria [1].

[2] The crossover point is that point on the trajectory at which the probability of being connected to the new base station is equal to that of being connected to the old base station [8].

(in decibels) from the new base station exceeds that from the current base station by a hysteresis level.

A drawback of signal-strength hysteresis algorithms is that they allow unnecessary handoffs in regions, where signals from both stations are strong [8]. One approach to prevent such handoffs has been to extend the hysteresis algorithm so that a handoff is not made in situations where the signal strength from the current base station is adequate [9], i.e., larger than some level needed for satisfactory performance. Our approach is to design an optimal algorithm based on a measure of call quality that incorporates this concept of an adequate signal strength. We use a binary-valued measure of call quality. Call quality is good or bad, depending on whether the signal strength is above or below a fixed threshold. It not difficult to identify such a threshold above which the subscriber will have satisfactory performance, but below which the distortion becomes unacceptable. This is especially true in digital systems, where the threshold level could represent the point at which reliable error correction is no longer possible. We refer to the event that the signal strength falls below the threshold as a *service failure*.[3]

Thus, in our definition, an optimal handoff algorithm is one that gives the best tradeoff between the expected number[4] of service failures and expected number of handoffs on a given mobile trajectory. The tradeoff problem can be posed in two ways: a variational and Bayes formulation. We establish that the solutions to the variational and Bayes formulations have the same structure (which is akin to the detection theory result that the Neyman–Pearson and Bayes tests are both likelihood ratio tests [10]).

The Bayes formulation of the handoff problem is amenable to *dynamic programming* (DP) arguments. We pose the problem as a finite-horizon[5] DP problem and obtain the optimal solution through a set of recursive equations. That optimal solution is complicated and nonstationary, and it requires prior knowledge of the mobile's exact trajectory. However, a simple, *locally optimal* algorithm can be derived from the DP solution. The locally optimal algorithm is a threshold algorithm (as is the signal-strength hysteresis algorithm), but it is not prone to unnecessary handoffs. Furthermore, it can be designed to be independent of the location of the mobile. We study this locally optimal algorithm in detail.

The rest of this paper is organized as follows. In Section II, we set up the handoff problem as an optimum tradeoff problem. In Section III, we present the DP solution. In Section IV, we discuss the locally optimal algorithm design. Section V provides detailed numerical results comparing the locally optimal algorithm to existing algorithms. The comparisons are done on the basis of the tradeoff between the expected number of service failures and handoffs. A discussion on adapting the locally optimal test to changing environments is also included. Conclusions are presented in Section VI.

[3] Occurrence of service failure does not imply that the call is dropped because we assume that the threshold level below which a service failure is declared may be greater than the level below which the call is lost.

[4] We assume a discrete-time model for the signal-strength measurements.

[5] This is in contrast with formulations of [11] and [12] that employ infinite-horizon discounted-cost models.

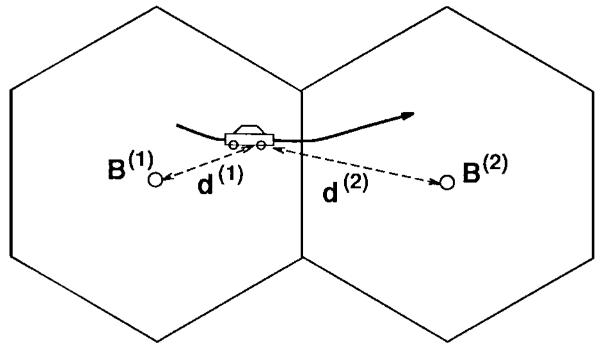

Fig. 1. A possible mobile trajectory.

II. Preliminaries

We assume that only two base stations, say $B^{(1)}$ and $B^{(2)}$, are involved in the handoff, i.e., we consider only that portion of the trajectory on which the signals received from base stations $B^{(1)}$ and $B^{(2)}$ are the strongest. Without loss of generality, assume that the mobile is moving on a trajectory going from cell 1 to cell 2, moving away from $B^{(1)}$ and approaching $B^{(2)}$. Let $d^{(i)}$ denote the distance of the mobile to $B^{(i)}, \ i = 1, 2$ (see Fig. 1).

We assume that the mobile measures the signal strength from each base station and sends sampled values to the operative base station. Furthermore, in our analysis, we assume that handoff decisions are based solely on these measurements. However, the analysis is easily extended to the case, where handoff decisions are also based on signal-strength measurements made at the base stations.

The measured signal strength has three components: path loss, which decays with distance from the base station, large-scale fluctuations (*shadow fading*), and small-scale fluctuations (*multipath fading*). We assume that the received signal is passed through a low-pass filter to average out the small-scale fluctuations. The reason for the averaging is two-fold. First, it is impractical to design handoff algorithms that respond to the small-scale fluctuations. Second, these fluctuations are generally compensated for by diversity combining and interleaving.

We assume that the signal strength is measured in decibels relative to some reference power level. (A convenient reference level, used in Section V, is the minimum signal strength required for satisfactory performance.) The signal strength $X^{(i)}$ received from base $B^{(i)}$ at distance $d^{(i)}$ (after low-pass filtering) can be written as

$$X^{(i)}(d^{(i)}) = \mu - \eta \log d^{(i)} + Z^{(i)}(d^{(i)}) \text{ dB}, \qquad i = 1, 2.$$

Parameters μ and η account for path loss, μ depends on transmitted power at the base station, and η is the path-loss exponent. The term $Z^{(i)}$ is the shadow fading component, which is accurately modeled (in decibels) as a zero-mean stationary Gaussian random process [13]. This model is referred to as the *lognormal* fading model. For spatial dependence, we assume a first-order autoregressive (AR) structure, which implies that the autocorrelation function of $Z^{(i)}$ is given by

$$\mathrm{E}[Z^{(i)}(d)Z^{(i)}(d+\delta)] = \sigma^2 \exp\left(-\frac{|\delta|}{d_0}\right)$$

where d_0 is the *correlation distance* and σ^2 is the variance of the shadow fading process.[6] This AR model has been proposed by Gudmundson [14] on the basis of experimental results. The spatial correlation is assumed to be *isotropic*, i.e., the correlation of the fading process at two locations depends only on the distance between the two locations. Furthermore, the processes $Z^{(1)}$ and $Z^{(2)}$ are independent.

Samples of the signal-strength measurements are relayed to the operative base station. Let the sampling time be t_s and assume that the mobile is moving at a constant velocity v. Then, with a slight abuse of notation, we get the following discrete-time model for the received signal strength:

$$X_k^{(i)} = \mu - \eta \log d_k^{(i)} + Z_k^{(i)} \tag{1}$$

where $d_k^{(i)}$ is the distance to base station $B^{(i)}$ at the kth sampling instant and $Z_k^{(i)}$ is kth sample of $Z^{(i)}$. The values $d_k^{(i)}$ are, in general, not equally spaced, but due to the isotropic nature of the fading process, consecutive samples of $Z^{(i)}$ have the same correlation if speed is fixed. If speed and direction are constant, the process $Z_k^{(i)}$ is a zero-mean Gaussian discrete-time AR process with autocorrelation function

$$\begin{aligned}\mathrm{E}[Z_k^{(i)} Z_{k+m}^{(i)}] &= \sigma^2 \exp\left(-\frac{|mvt_s|}{d_0}\right) = \sigma^2 \exp\left(-\frac{|md_s|}{d_0}\right)\\ &= \sigma^2 a^{|m|}\end{aligned}$$

where $d_s = vt_s$ is the *sampling distance* and $a = e^{-d_s/d_0}$ is the *correlation coefficient* of the discrete-time fading process.

Suppose there are a total of n time steps $k = 1, 2, \cdots, n$ on the portion of the mobile's trajectory that involves $B^{(1)}$ and $B^{(2)}$. Let B_k denote the index of the operative base station at time k (i.e., $B_k = i$ when the mobile is communicating with $B^{(i)}$) and B_k^c denote the other base station. A handoff decision is made during each sampling interval. The decision U_k at time k can be based on all signal-strength measurements up to time k. The decision variable U_k takes on two values. If $U_k = 1$, a handoff is made resulting in $B_{k+1} = B_k^c$. If $U_k = 0$, no handoff is made and $B_{k+1} = B_k$. Let I_k denote all the information available for decisionmaking at time k, then $U_k = \phi_k(I_k)$, where ϕ_k is the decision function at time k.

Handoff algorithm design involves choosing the handoff decision functions ϕ_k at times $k = 1, 2, \cdots, n-1$. Let Δ denote the minimum level of signal strength required for satisfactory service.[7] Furthermore, let N_{SF} and N_{H} denote the total number of service failures and number of handoffs from time one to n, respectively. Then

$$\mathrm{E}[N_{\mathrm{SF}}] = \mathrm{E}\left[\sum_{k=1}^{n} \mathbb{1}_{\{X_k^{(B_k)} < \Delta\}}\right] = \sum_{k=1}^{n} \mathrm{P}\{X_k^{(B_k)} < \Delta\}$$

and

$$\mathrm{E}[N_{\mathrm{H}}] = \mathrm{E}\left[\sum_{k=1}^{n-1} \mathbb{1}_{\{U_k = 1\}}\right] = \sum_{k=1}^{n-1} \mathrm{P}\{U_k = 1\}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

[6] Parameters μ, η, d_0, and σ could, in general, be different for the signals from the two base stations.

[7] If the interference levels for the two base stations are not the same, Δ will be different for the signals $X^{(1)}$ and $X^{(2)}$.

An optimal handoff algorithm (policy) is the set of decision functions $\phi = (\phi_1, \phi_2, \cdots, \phi_{n-1})$, which provides the best tradeoff between the $\mathrm{E}[N_{\mathrm{SF}}]$ and $\mathrm{E}[N_{\mathrm{H}}]$. This optimum tradeoff problem can be posed in two ways:

1) variational formulation

$$\min_{\phi} \ \mathrm{E}[N_{\mathrm{H}}] \text{ subject to } \mathrm{E}[N_{\mathrm{SF}}] \le \alpha \tag{2}$$

where α is a control parameter;

2) Bayes formulation

$$\min_{\phi} \ c\,\mathrm{E}[N_{\mathrm{H}}] + \mathrm{E}[N_{\mathrm{SF}}] \tag{3}$$

where $c > 0$ is a tradeoff parameter.

This kind of tradeoff commonly arises in detection problems [10], [15], where the solution to the variational problem is usually also a Bayes solution for an appropriately chosen tradeoff parameter. Theorem 1 of Section III shows that the same is true for the handoff problem. Hence, we can find the optimum tradeoff curve $(\mathrm{E}[N_{\mathrm{H}}], \mathrm{E}[N_{\mathrm{SF}}])$ for both problems by solving the Bayes problem for various values of c. Parameter c may be interpreted as the relative cost of handoffs versus service failures. This interpretation is especially useful for adapting the handoff algorithm to changing environments as we shall see in Section V. We focus now on the solution to the Bayes problem.

III. Dynamic Programming Solution

Dynamic programming allows optimization of the total cost along a state trajectory of a discrete-time dynamical system that has a stepwise additive-cost criterion and, conditioned on the state, stepwise independent-noise statistics (see [16], p. 10). Here, we cast the handoff problem in that form.

For the handoff problem, the state S_k at time k consists of the triple $(X_k^{(1)}, X_k^{(2)}, B_k)$. From (1) and the definition of B_k, we get the following update equation for S_k:

$$\begin{aligned}S_{k+1} &= \begin{bmatrix} X_{k+1}^{(1)} \\ X_{k+1}^{(2)} \\ B_{k+1}\end{bmatrix} = f(S_k, U_k, W_k)\\ &= \begin{bmatrix} X_k^{(1)} - \eta \log d_{k+1}^{(1)} + \eta \log d_k^{(1)} + W_k^{(1)} \\ X_k^{(2)} - \eta \log d_{k+1}^{(2)} + \eta \log d_k^{(2)} + W_k^{(2)} \\ B_k \mathbb{1}_{\{U_k=0\}} + B_k^c \mathbb{1}_{\{U_k=1\}}\end{bmatrix}\end{aligned}$$

where $W_k^{(i)} = Z_{k+1}^{(i)} - Z_k^{(i)}$ is the change in the fading process and $W_k = [W_k^{(1)}, W_k^{(2)}]$. The update equation $S_{k+1} = f(S_k, U_k, W_k)$ constitutes a discrete-time dynamical model for our system (with exogenous input W_k and control input U_k).

Given the first-order AR model for processes $\{Z_k^{(1)}\}$ and $\{Z_k^{(2)}\}$, the noise variables W_k have the required independence structure. In particular, it is easy to show that $W_{k+1}^{(i)}$ is independent of $\{W_1^{(i)}, \cdots, W_k^{(i)}\}$, given S_k.

Finally, the cost criterion as defined in (3) is additive over time. In particular, if we define

$$g_k(S_k, U_k) = c\mathbb{1}_{\{U_k=1\}} + \mathbb{1}_{\{X_k^{(B_k)} < \Delta\}}, \qquad 1 \le k < n \tag{4}$$

and

$$g_n(S_n) = \mathbb{1}_{\{X_n^{(B_n)} < \Delta\}}$$

then a Bayes optimal handoff algorithm minimizes

$$\mathrm{E}\,[g_n(S_n)] + \sum_{k=1}^{n-1} \mathrm{E}\,[g_k(S_k, U_k)].$$

Remark 1: The cost structure in (4) does not take into account the possibility that the call may be terminated at some time k between one and n. If ρ is the termination (hangup) probability [11] for each time step and if a geometric distribution is assumed for the call duration, then each cost term g_k should be multiplied by $(1-\rho)^k$. Thus, if a good estimate of ρ is available, it can easily be incorporated into the cost structure without any fundamental changes in the analysis.

The DP solution is obtained recursively as follows. Let the expected cost-to-go at time k (due to all the decisions up to time k) be denoted by J_k. Due to the conditionally independent noise statistics, J_k is a function only of the state S_k. The expected cost of the Bayes optimal handoff policy over the entire trajectory is simply $J_1(S_1)$, and optimal handoff decision functions are obtained by solving the following set of recursive equations (see [16] for an explanation of the DP technique):

$$J_n(X_n^{(1)}, X_n^{(2)}, B_n) = \mathbb{1}_{\{X_n^{(B_n)} < \Delta\}} \tag{5}$$

$$\begin{aligned} J_{n-1}&(X_{n-1}^{(1)}, X_{n-1}^{(2)}, B_{n-1}) \\ &= \min_{\phi_{n-1}} \mathrm{E}_{W_{n-1}|S_{n-1}}[g_n(S_n) + g_{n-1}(S_{n-1}, U_{n-1})] \\ &= \mathbb{1}_{\{X_{n-1}^{(B_{n-1})} < \Delta\}} + \min\,[\mathrm{P}\{X_n^{(B_{n-1})} < \Delta \mid X_{n-1}^{(B_{n-1})}\},\ c \\ &\quad + \mathrm{P}\{X_n^{(B_{n-1}^c)} < \Delta \mid X_{n-1}^{(B_{n-1}^c)}\}] \end{aligned} \tag{6}$$

and for $k = n-2, n-3, \cdots, 1$

$$\begin{aligned} J_k&(X_k^{(1)}, X_k^{(2)}, B_k) \\ &= \min_{\phi_k} \mathrm{E}_{W_k|S_k}(J_{k+1}(S_{k+1}) + g_k(S_k, U_k)) = \mathbb{1}_{\{X_k^{(B_k)} < \Delta\}} \\ &\quad + \min\,(\mathrm{E}\{J_{k+1}(X_{k+1}^{(1)}, X_{k+1}^{(2)}, B_k) \mid X_k^{(1)}, X_k^{(2)}\},\ c \\ &\quad + \mathrm{E}\{J_{k+1}(X_{k+1}^{(1)}, X_{k+1}^{(2)}, B_k^c) \mid X_k^{(1)}, X_k^{(2)}\}). \end{aligned} \tag{7}$$

Note that for each k, the optimum decision function $\phi_k^\star$ depends only on the state S_k and not on any past signal-strength measurements. These optimum decision functions are described by

$$\begin{aligned} &\mathrm{P}\{X_n^{(B_{n-1}^c)} < \Delta \mid X_{n-1}^{(B_{n-1}^c)}\} + c \\ &\quad \underset{U_{n-1}=1}{\overset{U_{n-1}=0}{\gtrless}} \mathrm{P}\{X_n^{(B_{n-1})} < \Delta \mid X_{n-1}^{(B_{n-1})}\} \end{aligned} \tag{8}$$

and for $k = 1, 2, \cdots, n-2$

$$\begin{aligned} &\mathrm{E}\{J_{k+1}^{(B_k^c)}(X_{k+1}^{(1)}, X_{k+1}^{(2)}) \mid X_k^{(1)}, X_k^{(2)}\} \\ &\quad + c \underset{U_k=1}{\overset{U_k=0}{\gtrless}} \mathrm{E}\{J_{k+1}^{(B_k)}(X_{k+1}^{(1)}, X_{k+1}^{(2)}) \mid X_k^{(1)}, X_k^{(2)}\}. \end{aligned} \tag{9}$$

The decision functions $(\phi_1^\star, \phi_2^\star, \cdots, \phi_{n-1}^\star)$ described in (8) and (9) constitute a Bayes optimal handoff algorithm $\phi^\star$ for tradeoff parameter c. The following theorem, whose proof is given in the Appendix, addresses the solution to the variational formulation of (2).

Theorem 1: An optimal handoff algorithm for the variational formulation is a Bayes solution for an appropriately chosen value of the tradeoff parameter c.

The DP equations can be solved numerically to obtain the optimum handoff policy $\phi^\star$, however, the solution can be quite complex and nonstationary. Furthermore, the computation of $\phi^\star$ relies on prior knowledge of the trajectory of the mobile. These considerations imply that the DP solution is impractical. The tradeoff curve for the DP solution can, however, be used as a theoretical benchmark in the comparison of other suboptimal algorithms. In the next section, we discuss the design of a simple stationary locally optimal handoff algorithm that is based on the DP solution.

IV. Locally Optimal Algorithm

Solutions (8) and (9) correctly indicate that the globally optimal strategy at a particular location depends on the future trajectory. That unreasonable requirement suggests that the problem should be reformulated specifically to ignore the future trajectory. A locally optimal solution to the Bayes problem may be obtained by restricting the trajectory under consideration to the points k and $k+1$. That is, we ignore the consequences of a handoff decision at time k at times beyond $k+2$ and base the decision on all available information up to time k. Restricting (8) and (9) to $n = 2$ yields decision rules ϕ_k that select the best tradeoff between the cost of a handoff and the probability that $X_{k+1}^{(B_{k+1})}$ falls below Δ, given the information I_k. Hence, the locally optimal decision function ϕ_k^{lo} at time k has the structure

$$\mathrm{P}\{X_{k+1}^{(B_k^c)} < \Delta \mid I_k\} + c \underset{U_k=1}{\overset{U_k=0}{\gtrless}} \mathrm{P}\{X_{k+1}^{(B_k)} < \Delta \mid I_k\}. \tag{10}$$

The local criterion gives rise to a solution that uses only local information.

Remark 2: Suppose we account for possible call termination in the design of the locally optimal test by multiplying each cost function g_k by the discount factor $(1-\rho)^k$. The only modification that results in the locally optimal decision function ϕ_k^{lo} is that the parameter c is replaced by $c(1-\rho)$. The structure of ϕ_k^{lo} remains unchanged. Furthermore, if ρ is small, ϕ_k^{lo} is virtually unaffected by the discount factor.

For the lognormal fading model that we have assumed, the conditional distribution of $X_{k+1}^{(i)}$ given $X_k^{(i)}$ is Gaussian, hence, the probabilities in (10) are entirely determined by the conditional means and variances

$$\mathrm{E}[X_{k+1}^{(i)} \mid I_k] = aX_k^{(i)} + (1-a)\,\mu - \eta \log\left[\frac{d_{k+1}^{(i)}}{\left(d_k^{(i)}\right)^a}\right] \tag{11}$$

$$\mathrm{Var}[X_{k+1}^{(i)} \mid I_k] = (1-a^2)\sigma^2. \tag{12}$$

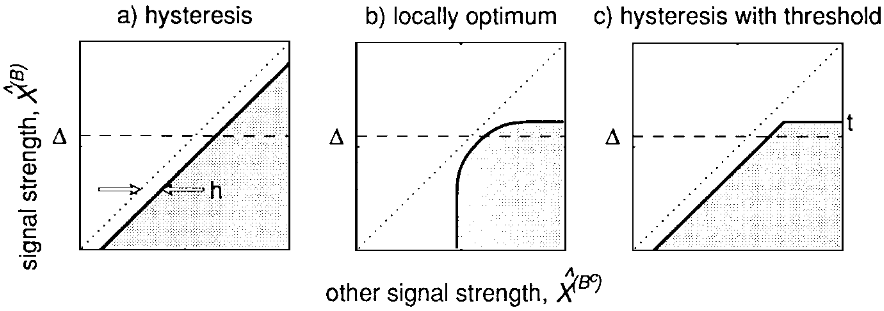

Fig. 2. Decision regions based on the predicted signal strength $\hat{X}^{(B)}$ of the current station and the other station $\hat{X}^{(B^c)}$, respectively, for three different handoff strategies. Decision in favor of handoff is made when the signal strengths lie inside the shaded area.

Note that the conditional variance does not depend on location information. Furthermore, if the sampling rate is high, the argument of the log function in (11) is close to one, which implies the conditional mean is also approximately independent of location information. However, with $d_k^{(i)}$ and $d_{k+1}^{(i)}$ unknown, we are, in practice, forced to use the best available estimate of (11), denoted $\hat{X}_{k+1}^{(i)}$, based on the available information I_k. The resulting test ϕ_k^{lo} is described by

$$Q\left[\frac{\Delta - \hat{X}_{k+1}^{(B_k^c)}}{\sigma\sqrt{1-a^2}}\right] + c \underset{U_k=1}{\overset{U_k=0}{\gtrless}} Q\left[\frac{\Delta - \hat{X}_{k+1}^{(B_k)}}{\sigma\sqrt{1-a^2}}\right] \tag{13}$$

where $Q(x) = 1/\sqrt{2\pi}\int_x^\infty \exp(-x^2/2)\,dx$. The correlation parameter $a = \exp(vt_s/d_0)$ is speed dependent and, in practice, it can be determined from velocity estimates. Several authors treat the problem of estimating statistics of the fading environment [17]–[20], and the algorithm suggested by (13) can clearly accommodate on-line estimates of v, d_0, and σ. This adaptation is discussed further in Section V.

Fig. 2 shows the decision regions of the conventional hysteresis algorithm, locally optimum test, and hysteresis-threshold algorithm presented by Zhang and Holtzman [9]. Comparing the hysteresis test and the locally optimum test, we see the locally optimum test can avoid unnecessary handoffs in situations where both signal strengths are above Δ. In systems designed to produce very few service failures, i.e., those with large fade margins, this could result in a significant saving in the number of handoffs. The locally optimum test also avoids bad handoffs when both signals are below Δ. However, in systems with large fade margins, that situation occurs very rarely.

An *ad hoc* scheme that emulates the behavior of the locally optimum test is the hysteresis-threshold algorithm [9], which performs the usual hysteresis test $\hat{X}^{(B^c)} \gtrless \hat{X}^{(B)} + h$, but only after first verifying that $\hat{X}^{(B)} < t$ with a suitably chosen threshold. In Section V, we see that the performance of a locally optimum test can be closely approximated by the hysteresis-threshold test, but the advantage in doing so is not clear. Unlike the locally optimum test, the hysteresis-threshold test requires two design parameters and is not easily adapted to changing environments.

V. Numerical Results

We simulate a scenario in which the mobile traverses the line leading a distance D from station 1 to station 2. The specific simulation parameters were:

$D = 2000$ m	station distance;
$\mu = 105$	station strength;
$\eta = 30$	path-loss exponent;
$\sigma^2 = 25$	fading process variance;
$d_s = 2, 5,$ and 10 m	sampling distances;
$d_0 = 30$ m	correlation distance;
$\Delta = 0$ dB	threshold of service failure.

These values are consistent with [9] and [14] and together yield a median signal strength of 15 dB (with respect to Δ) at the midpoint between stations. At a sampling period of $t_s = 0.5$ s, the sampling distances of 2, 5, and 10 m correspond to speeds of 14.4, 36, and 72 km/h.

50 000 realizations were used to estimate the performance of each strategy at each parameter setting. The same signal-strength predictor $\hat{X}_{k+1}^{(i)} = X_k^{(i)}$ was used in comparing decision strategies. The Bayes optimal solution was constructed by quantizing the decision regions (akin to Fig. 2) for ϕ_k^* and calculating the expectations in (5)–(7) by numerical integration.

Fig. 3 compares the performance of the locally optimum handoff algorithm with that of straightforward hysteresis and hysteresis-threshold tests in terms of the tradeoff between the number of handoffs and service failures. Compared to the simple hysteresis test, the locally optimum and threshold strategies consistently achieve fewer service failures for the same number of handoffs. The locally optimum test has, effectively, the same performance as the best of threshold tests. More precisely, the number of service failures for the locally optimum test can slightly exceed that of the hysteresis-threshold test, but, by varying a single parameter c, the locally optimum performance curve appears to trace close to the minimum $\min_{h,t} \mathrm{E}[N_{\mathrm{SF}}]$ of all threshold tests. We have no strategy to construct the best performance curve for all hysteresis-threshold tests. However, experimentation shows that the best hysteresis-threshold tests employ small hysteresis values (such as 0, 1, and 2 dB used in the figure) to which their performance is somewhat insensitive.

Fig. 4 compares performance of the locally optimum decision rule at three sampling distances corresponding to different speeds. Due to spatial correlation, short-sampling distances allow better signal-strength prediction and, thus, better handoff decisions and fewer service failures. Dashed lines show how the operating point changes if the tradeoff parameter remains

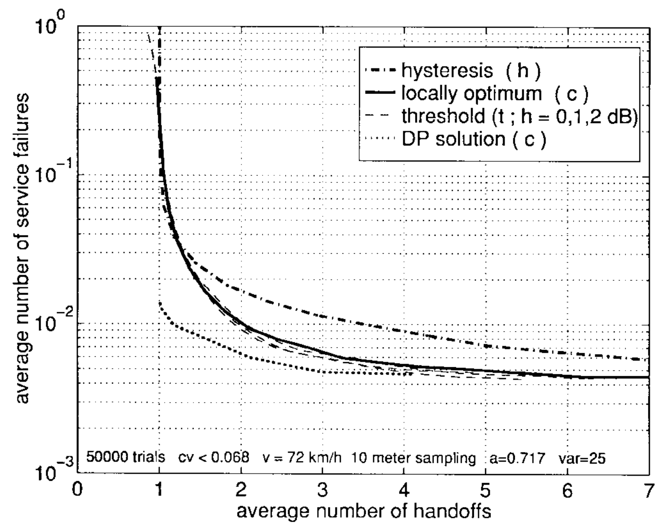

Fig. 3. Performance comparison. We compare the tradeoff between handoffs and service failures for three strategies to each other and to the best achievable (dynamic programming solution). Each curve is parameterized by the variable in parenthesis. With a single parameter, the locally optimum test follows the best performance that is achievable using two parameter hysteresis-threshold tests.

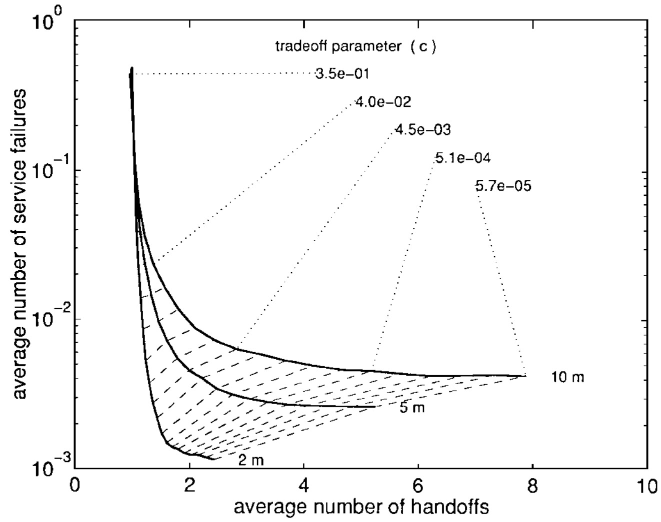

Fig. 4. Tradeoff curve parametrization. Performance of the locally optimum test is shown at sampling distances corresponding to mobile speeds 14.4, 36, and 72 km/h (assuming $t_s = 0.5$ s). Dashed lines connect points of the equal tradeoff parameter. Unlabeled tradeoff values are spaced logarithmically between indicated values.

fixed, while the cellular environment (e.g., speed) changes. In making (13) adaptive, we may clearly use on-line estimates of the shadow fading variance and signal-strength correlation, but is it also necessary to adapt the tradeoff parameter c? Fig. 4 shows that if the tradeoff parameter remains fixed, then the "knee" of the operating curve at slow speeds transforms nearly to the knee of the curve at high speeds. Thus, an acceptable test can be had by choosing a fixed value for the tradeoff parameter. This compares favorably with approaches to adaptive hysteresis tests that require separate analysis at each speed in order to determine the optimal hysteresis value [17].

Fig. 5 shows performance curves of the simple hysteresis test as a function of sampling distance. Note, in comparison

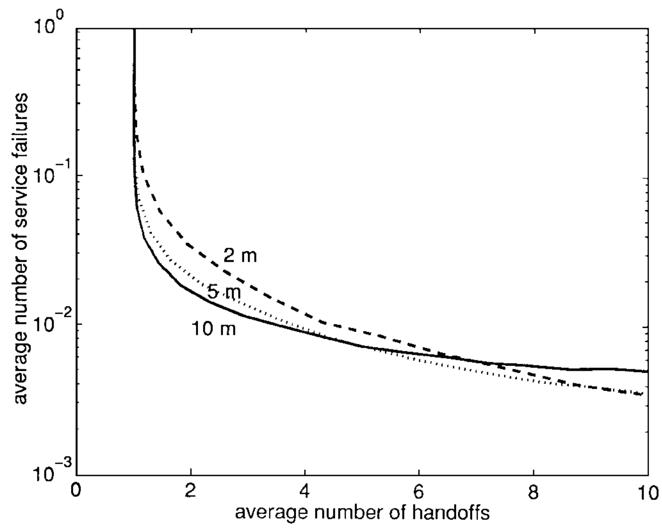

Fig. 5. The hysteresis algorithm does not adapt to speed. Performance of the hysteresis test is shown at sampling distances corresponding to mobile speeds 14.4, 36, and 72 km/h (assuming $t_s = 0.5$ s). Compared with the locally optimal test (Fig. 4), the achievable performance of hysteresis tests is relatively unchanged at different mobile speeds.

to Fig. 4, that hysteresis performance curves are relatively unchanged at different sampling distances. A speed-adaptive rule could ideally move the operating point to any point on the particular curve specified by the current mobile speed. But even with such a rule, hysteresis tests cannot reach the same points in the $(\mathrm{E}[N_{\mathrm{H}}], \mathrm{E}[N_{\mathrm{SF}}])$ plane as the locally optimal test can. Hysteresis tests cannot take advantage of the increased predictability of the signal strength at slow speeds.

VI. CONCLUSIONS

We introduced a new call-quality criterion to balance against the number of handoffs in designing an optimal handoff strategy and showed that the Bayes and variational formulations of the resulting optimization are equivalent. The optimal decision rules may be found by dynamic programming, but are too costly to implement and depend on prior knowledge of the trajectory of the mobile.

A locally optimal solution gives rise to a hysteresis test that compares probabilities rather than signal strengths. This locally optimal test naturally has the property of preventing handoffs when the signal from the operative base station is strong and allowing handoffs when that signal is weak. The locally optimum test compares favorably in performance with simple hysteresis tests and is competitive with hysteresis-threshold tests. For the locally optimum strategy, it is immediately clear how to incorporate on-line parameter estimates to obtain an algorithm that responds to changes in the propagation environment. This natural adaptability is the principle advantage of the locally optimal handoff test over current approaches.

APPENDIX

Proof of Theorem 1: Let $J(\phi, c)$ denote the Bayes cost for a handoff policy ϕ, i.e.,

$$J(\phi, c) = c\mathbf{E}[N_{\mathrm{H}}](\phi) + \mathbf{E}[N_{\mathrm{SF}}](\phi)$$

where $\mathrm{E}[N_{\mathrm{H}}]$ and $\mathrm{E}[N_{\mathrm{SF}}]$ are written explicitly as functions of ϕ. The Bayes optimal solution $\phi_c^\star$ has a cost $J^\star(c)$ that is given by

$$J^\star(c) = \min_\phi J(\phi, c) = J(\phi_c^\star, c).$$

The following lemma gives an important property of $J^\star(c)$.

Lemma 1: $J^\star(c)$ is a concave continuous function of c on $(0, \infty)$.

Proof: The Bayes optimal solution is obtained recursively from (5) to (9). In particular, $J^\star(c) = J_1(S_1)$. In the following, we write the cost-to-go functions explicitly as functions of c, i.e., $J_k(S_k, c)$.

The function $J_N(S_N, c)$ is trivially a concave function of c. Now, suppose $J_{k+1}(S_{k+1}, c)$ is a concave function of c. Then, from (6), it follows that $J_k(S_k, c)$ is a concave function of c as well since $J_k(S_k, c)$ is expressed as the sum of a constant and the minimum of two concave functions of c. Thus, by induction, $J_1(S_1, c) = J^\star(c)$ is also a concave (hence continuous) function of c on $(0, \infty)$.

Now, for any fixed policy ϕ, the Bayes cost $J(\phi, c)$ is a straight line as a function of c with slope $E[N_{\mathrm{H}}](\phi)$ and intercept $E[N_{\mathrm{SF}}](\phi)$. Furthermore, this line must lie above the concave curve $J^\star(c)$. The line corresponding to the variational solution ϕ_α is one that has an intercept less than or equal to α and the minimum possible slope. It is clear that such a line must pass through a point (or points) on curve $J^\star(c)$. Let the ordinate of one such point be c_α. Then, the Bayes solution for c_α is a solution to the variational problem.

ACKNOWLEDGMENT

The authors thank R. Vijayan for helpful discussions and R. Rezaiifar for providing an early copy of [11] that gives the call-termination interpretation for the discounted cost structure.

REFERENCES

[1] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, vol. 29, no. 11, pp. 42–46, Nov. 1991.

[2] B. Gudmundson and O. Grimlund, "Handoff in microcellular based personal telephone systems," in *Third Generation Wireless Information Networks*, S. Nanda and D. J. Goodman, Eds. Boston, MA: Kluwer, 1992, pp. 187–203.

[3] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, 1991.

[4] K. J. Bye, "Handover criteria and control in cellular and microcellular systems," in *5th Int. Conf. Mobile Radio Personal Commun.*, Dec. 1989, pp. 94–98.

[5] E. A. Frech, "Cellular models and hand-off criteria," in *Proc. IEEE 39th Veh. Technol. Conf.*, 1989, pp. 128–135.

[6] S. T. S. Chia and R. J. Warburton, "Handover criteria for city microcellular radio systems," in *Proc. IEEE 40th Veh. Technol. Conf.*, 1990, pp. 276–281.

[7] A. Murase, I. C. Symington, and E. Green, "Handover criterion for macro and microcellular systems," in *Proc. IEEE 41st Veh. Technol. Conf.*, 1991, pp. 524–530.

[8] R. Vijayan and J. M. Holtzman, "A model for analyzing handoff algorithms," *IEEE Trans. Veh. Technol.*, vol. 42, no. 3, pp. 351–356, 1993.

[9] N. Zhang and J. M. Holtzman, "Analysis of handoff algorithms using both absolute and relative measurements," in *Proc. IEEE 44th Veh. Technol. Conf.*, 1994, pp. 82–86.

[10] H. V. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 1988.

[11] R. Rezaiifar, A. M. Makowski, and S. P. Kumar, "Stochastic control of handoffs in cellular networks," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1348–1362, 1995.

[12] M. Asawa and W. E. Stark, "A framework for optimal scheduling of handoffs in wireless networks," in *Proc. IEEE Globecom '94*, Nov. 1994, pp. 1669–1673.

[13] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[14] M. Gudmundson, "Analysis of handover algorithms," in *Proc. IEEE 41st Veh. Technol. Conf.*, 1991, pp. 537–542.

[15] A. N. Shiryayev, *Optimal Stopping Rules.* New York: Springer-Verlag, 1978.

[16] D. Bertsekas, *Dynamic Programming.* Upper Saddle River, NJ: Prentice-Hall, 1987.

[17] M. D. Austin and G. L. Stüber, "Velocity adaptive handoff algorithms for microcellular systems," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 549–561, 1994.

[18] W. C. Y. Lee and Y. S. Yeh, "On the estimation of the second-order statistics of log normal fading in mobile radio environment," *IEEE Trans. Commun.*, vol. 22, pp. 869–873, June 1974.

[19] P. Dassanayake, "Dynamic adjustment of propagation dependent parameters in handover algorithms," in *Proc. IEEE 44th Veh. Technol. Conf.*, 1994, pp. 73–76.

[20] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and shadowing," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 439–446, 1994.

Venugopal V. Veeravalli (S'84–M'93) received the B.Tech. degree (Silver Medal Honors) in 1985 from the Indian Institute of Technology, Bombay, India, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, all in electrical engineering.

From 1992 to 1993, he was a Post-Doctoral Fellow in the Division of Applied Sciences, Harvard University, Cambridge, MA. In August 1996, he joined the faculty of the School of Electrical Engineering, Cornell University, Ithaca, NY, as an Assistant Professor. He has also held academic positions at CUNY and Rice University, Houston, TX. His research interests are in the areas of wireless communications, detection and estimation theory, and information theory. His current topics of research include capacity analysis for cellular CDMA systems, sequential acquisition techniques for CDMA systems, handoff and power control algorithms for wireless networks, and decentralized dynamic decision making.

Dr. Veeravalli is a corecipient of the 1996 IEEE Browder J. Thompson Best Paper Award.

Owen E. Kelly (S'91–M'96) received the B.A.Sc. degree in systems design engineering in 1989 from the University of Waterloo, Ont., Canada, and the M.S. and Ph.D. degrees in electrical engineering in 1994 and 1996 from Rice University, Houston, TX.

He is currently a Post-Doctoral Fellow at the Wireless Information Network Laboratory, Rutgers University, Piscataway, NJ, working on mobile location estimation and simulation of large wireless networks. His research interests are in the areas of statistical modeling and decision problems and the application of related techniques in the field of wireless communications.