Controlled Flooding Search In a Large Network

Nicholas B. Chang, Mingyan Liu
Electrical Engineering and Computer Science Department
University of Michigan, Ann Arbor
{changn, mingyan}@eecs.umich.edu

***Abstract*— In this paper we consider the problem of searching for a node or an object (i.e., piece of data, file, etc.) in a large network. Applications of this problem include searching for a destination node in a mobile ad hoc network, querying for a piece of desired data in a wireless sensor network, and searching for a shared file in an unstructured peer-to-peer network. We consider the class of controlled flooding search strategies where query/search packets are broadcast and propagated in the network until a preset TTL (time-to-live) value carried in the packet expires. Every unsuccessful search attempt, signified by a timeout at the origin of the search, results in an increased TTL value (i.e., larger search area) and the same process is repeated until the object is found. The primary goal of this study is to find search strategies (i.e., sequences of TTL values) that will minimize the cost of such searches associated with packet transmissions. Assuming that the probability distribution of the object location is not known *a priori*, we derive search strategies that minimize the search cost in the worst-case, via a performance measure in the form of the competitive ratio between the average search cost of a strategy and that of an omniscient observer. This ratio is shown in prior work to be asymptotically (as the network size grows to infinity) lower bounded by 4 among all deterministic search strategies. In this paper we show that by using randomized strategies (i.e., successive TTL values are chosen from certain probability distributions rather than deterministic values), this ratio is asymptotically lower bounded by e. We derive an optimal strategy that achieves this lower bound, and discuss its performance under other criteria. We further introduce a class of randomized strategies that are sub-optimal but potentially more useful in practice.**

***Index Terms*— Query and search, TTL, controlled flooding search, wireless networks, randomized strategy, best worst-case performance, competitive ratio**

I. Introduction

In this paper we consider the problem of searching for a node or an object (e.g., piece of data, file, etc.) in a large wireless network. A prime example is data query in a wireless sensor network, where different sensing data is distributed among a large number of sensor nodes [1]. Search has also been extensively used in mobile ad hoc networks, including searching for a destination node by a source node in the route establishment procedure of an ad hoc routing protocol (e.g., [2]), searching for a multicast group by a node looking to join the group (e.g., [3]), and locating one or multiple servers by a node requesting distributed services (e.g., [4]). Search is also widely used in peer-to-peer (P2P) networks.

A variety of mechanisms may be used to locate a node in a large network. For instance, a centralized directory service, which is periodically updated, can be established from which location information may be obtained. One can also use the decentralized random walk based search, where the querier sends out a query packet to be forwarded in some random fashion, e.g., random walks or controlled walks such that the propagation of the packet follows a consistent direction, until it hits the search target [1].

In this paper we focus on a widely used search mechanism known as the TTL-based controlled flooding of query packets. Under this scheme the query/search packet is broadcast and propagated in the network. A preset TTL (time-to-live) value is carried in the packet and every time the packet is relayed the TTL value is decremented. This continues until TTL reaches zero and the propagation stops. Therefore the extent/area of the search is controlled by the TTL value. If the target is located within this area, it will reply with the queried information. Otherwise, the origin of the search will eventually time out and initiate another round of search covering a bigger area using a larger TTL value. This continues until either the object is found or the querier gives up. Consequently the performance of a search strategy is determined by the sequence of TTL values used.

Our primary goal is to derive controlled flooding search strategies, i.e., sequences of TTL values, that minimize the cost of such searches in terms of energy consumption (i.e., the amount of packet transmission/reception) [1]. We will mainly limit our analysis to the case of searching for a single target, which is assumed to exist in the network. It will be seen later that our results apply to the more general case of searching for multiple objects. For the rest of our discussion we will use the term *object* to indicate the target of a search, be it a node, a piece of data or a file. We measure the position of an object by its distance to the source originating the search. We will use the term *location* of an object to indicate both the actual position of the search target and the minimum TTL value required to locate this object.

When the probability distribution of the object location is known *a priori*, search strategies that minimize the expected search cost can be obtained via a dynamic programming formulation [5]. When the distribution of the object location

This work is supported by NSF grants ANI-0238035, CCR-0329715, ONR grant N00014-03-1-0232, and through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

[1] We will not explicitly consider the response time of a search strategy in this paper, as within the class of controlled flooding search the fastest search is to flood the entire network.

is not known *a priori*, one may evaluate the effectiveness of a strategy by its worst case performance. In [6] such a criterion, in the form of the competitive ratio (or worst-case cost ratio) between the expected cost of a given strategy and that of an omniscient observer, was used. It was shown under a linear cost model (to be precisely defined in the next section) that the best worst-case search strategy among all *fixed strategies* is the California Split search algorithm, which achieves a competitive ratio of 4 (also the lower bound on all fixed strategies). In [5] we showed that to minimize this ratio, the best strategies are *randomized* strategies that consist of sequence of random variables instead of deterministic values. In particular, [5] introduced a class of *uniformly randomized strategies* and showed that within this class the best strategy achieves a competitive ratio of approximately 2.9142.

In this paper we show that for a much more general class of cost models, the best worst-case strategy among all fixed and random strategies achieves a competitive ratio of e. We derive an optimal randomized strategy that attains this ratio and discuss how it can be adjusted to account for alternative performance criteria. We also establish an equivalence relationship between TTL sequences under different cost functions. This allows us to derive good randomized strategies for general cost functions based on the optimal uniformly randomized strategy derived for linear cost functions. These are sub-optimal strategies, but are simple to implement and of practical value.

The rest of the paper is organized as follows. Sections II and III present the network model and the performance objectives under consideration. In Section IV we derive the optimal strategy among all random and non-random strategies. We examine a few alternative performance measures in Section V. We establish a mapping between linear and more general cost functions in Section VI. Using this result in Section VII, we investigate a number of sub-optimal search strategies in the interest that these may be more practical and easier to implement in many cases. Section VIII concludes the paper.

II. Network Model

Within the context of TTL-based controlled flooding search, the distance between two nodes is measured in number of *hops*, assuming that the network is connected. Two nodes being one hop away means they can reach each other in one transmission. We will assume that a query with TTL value k will reach all nodes within k hops of the originating node before the next round of search starts. This is a simplification, but nevertheless allows us to reveal fundamental features of the problem and obtain insights. We denote by L the minimum TTL value required to search every node within the network, and will also refer to L as the *dimension* or *size* of the network. Since we have assumed that the object exists, using a TTL value of L will locate the object with probability 1.

A search strategy $\mathbf{u}$ is a TTL sequence of certain length N, $\mathbf{u} = [u_1, u_2, \cdots, u_N]$. It can be either fixed/deterministic where $u_i, i = 1, \cdots, N$, are deterministic values, or random where u_i are drawn from probability distributions. For a fixed strategy we assume that $\mathbf{u}$ is an increasing sequence, i.e., $u_i < u_{i+1}$ for $1 \leq i \leq N-1$. For randomized strategies, we assume all realizations are increasing sequences. In practice, it is natural to consider discrete (or integer-valued) policies. However, considering real-valued sequences can often reveal fundamental properties that are helpful in deriving optimal integer-valued strategies. In addition, real-valued strategies may also have practical applications, e.g., in ad hoc networks that use position information, flooding may be done within a real-valued physical distance (determined by the TTL) around the source. We therefore also consider continuous (or real-valued) strategies, denoted by $\mathbf{v}$, where $\mathbf{v} = [v_1, v_2, \cdots, v_N]$, and v_i is either a fixed or continuous random variable that takes any real value on $[1, \infty)$, for $1 \leq i \leq L$.

A strategy is *admissible* if it locates any object of finite location with probability 1. For a fixed strategy this implies $u_N = L$, and for a random strategy, this implies $Pr(u_i = L \text{ for some } 1 \leq i \leq N) = 1$. In the asymptotic case as $L \to \infty$, a strategy $\mathbf{u}$ is admissible if $\forall\ x \geq 1, Pr\left(u_n \geq x \text{ for some } n \in \mathbb{Z}^+\right) = 1$.

We let V denote the set of all real-valued admissible strategies (random or fixed). V^d denotes the set of all admissible real-valued deterministic strategies. U denotes the set of all integer-valued admissible strategies (random or fixed). Finally, U^d denotes the set of all admissible integer-valued deterministic strategies. Note that it is always true that $U^d \subset U \subset V$, and similarly $U^d \subset V^d \subset V$.

In a practical system, a variety of techniques may be used to reduce the number of query packets flowing in the network and to alleviate the *broadcast storm* problem [7]. In our analysis we will assume that a search with a TTL value of k will reach all neighbors that are k hops away from the originating node, and that the cost associated with this search is a function of k, denoted by $C(k)$. This cost may include the total number of transmissions, receptions, etc. Thus $C(k)$ is the abstraction of the nature of the underlying network and the specific broadcast schemes used.

It is important to note that in general a node receiving the search query will be unaware whether the object is found at another node in the same round (except perhaps when the object is found at one of its neighbors). Thus this node will continue decrementing the TTL value and passing on the search query. We can therefore regard the search cost as being *paid in advance*, i.e., the search cost for each round is determined by the TTL value and not by whether the object is located in that round. Two example cost functions are the linear cost and quadratic cost, defined as $C(k) = \alpha k$ and $C(k) = \alpha k^2$, respectively, for some constant $\alpha > 0$. When cost measures the number of transmissions, the first is a good model in a linear network with constant node density. The latter is a more reasonable model for a two-dimensional network, as the number of nodes reached (as well as the number of transmissions) in k hops has been shown to be on the order of k^2 [5], [6].

For real-valued sequences, we require that the cost function $C(v)$ be defined for all $v \in [1, \infty)$, while for integer-valued sequences we only require that the cost function be defined for positive integers. When the cost function is invertible, we use $C^{-1}(\cdot)$ to denote its inverse. We will adopt the natural assumption that $C(v_1) > C(v_2)$ if $v_1 > v_2$.

Both [6] and [5] considered only the linear cost function scenario where it is assumed that $C(k) = \alpha k$ for come constant α. In this paper we will consider a much broader class of cost functions defined as follows.

Definition 1: The function $C : [1, \infty) \to [C(1), \infty)$ belongs to the class $\mathbb{C}$ if $0 < C(1) < \infty$, $C(v)$ is increasing and differentiable (hence continuous), and $\lim_{v\to\infty} C(v) = \infty$.

We will use X to denote the minimum TTL value required to locate the object. We will also refer to X as the object *location*. As a result, an object location is an integer (real number) when discrete (continuous) strategies are considered. We denote the cumulative distribution of X by $F(x)$, where $F(x) = Pr(X \leq x)$. Similarly, the tail distribution of X is denoted by $\bar{F}(x) = 1 - F(x) = Pr(X > x)$. In the more general case of searching for k out of n objects, we can let X denote the location of the kth furthest object from the source. In this case the search process proceeds in exactly the same way as if searching for a single object with location X and terminates when all k objects have been found. Note that all $k - 1$ objects located closer to the source get a "free ride", i.e., they are automatically found either before or at the same time the kth furthest one is found. Therefore, without loss of generality we can assume there exists exactly one object in the network even though our results directly apply to searching for a subset of multiple objects.

III. Problem Formulation and Preliminaries

We adopt the following worst-case performance measure (a generalization of the one used in [6]):

$$\rho^{\mathbf{u}} = \sup_{\{p_X(x)\}} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \ , \tag{1}$$

where $J_X^{\mathbf{u}}$ denotes the expected search cost of using strategy $\mathbf{u}$ for object location X; $E[C(X)]$ is the expected search cost of an ideal omniscient observer who knows precisely the location (i.e., realization of X). The ratio between these two terms for a given X will be referred to as the (average) *cost ratio*. Meanwhile, $\{p_X(x)\}$ denotes the set of all probability mass functions of X such that $E[C(X)] < \infty$. We will only consider the case where the random vector $\mathbf{u}$ and X are mutually independent, as the distribution of X is not known *a priori*. Let $j_X^{\mathbf{u}}$ denote the search cost (a random variable in general) of using strategy $\mathbf{u}$ when object location is X. This can be written as:

$$j_X^{\mathbf{u}} = \sum_{u_i \in \mathbf{u}} C(u_i) I(X > u_{i-1}) \ , \tag{2}$$

Then $J_X^{\mathbf{u}}$ can be calculated as follows:

$$J_X^{\mathbf{u}} = E_{\mathbf{u}}\left[E_X\left[j_X^{\mathbf{u}}\right]\right] = E_{\mathbf{u}}\left[\sum_{u_i \in \mathbf{u}} C(u_i) Pr(X > u_{i-1})\right] \ , \tag{3}$$

where $u_0 = 0$ is assumed for all $\mathbf{u}$. Note that if $\mathbf{u}$ is deterministic then $J_X^{\mathbf{u}}$ is a single expectation with respect to X, whereas if $\mathbf{u}$ is random then $J_X^{\mathbf{u}}$ is the average over both X and $\mathbf{u}$. The corresponding objective is to find search strategies that minimize this ratio, with the best worst-case discrete strategy denoted by $\mathbf{u}^*$:

$$\rho^* = \inf_{\mathbf{u} \in U} \rho^{\mathbf{u}} = \inf_{\mathbf{u} \in U} \sup_{\{p_X(x)\}} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \ . \tag{4}$$

The worst-case cost ratio $\rho^{\mathbf{u}}$ can also be viewed as the *competitive ratio* with respect to an *oblivious adversary* [8] who knows the search strategy $\mathbf{u}$. We will use these two terms interchangeably. It should be mentioned that the quantity $\rho^{\mathbf{u}}$ has a slightly different meaning for deterministic and randomized strategies. When $\mathbf{u}$ is a fixed sequence, $J_X^{\mathbf{u}}$ is a single expectation with respect to X as seen in (3). In this case, for any given location the search cost of using $\mathbf{u}$, $J_X^{\mathbf{u}}$, is always within a factor $\rho^{\mathbf{u}}$ of the omniscient observer's cost. On the other hand, when $\mathbf{u}$ is random, then $\rho^{\mathbf{u}}$ only provides an upper bound on the *average* search cost but does not necessarily upper bound any particular realization of this cost. In this case, it is the *expected* search cost of $\mathbf{u}$ that is always within $\rho^{\mathbf{u}}$ of the cost of an omniscient observer. In Section V, we will present other performance measures in order to account for these differences.

For any continuous strategy, $\mathbf{v} \in V$, the worst-case cost ratio is defined similarly to (1): $\rho^{\mathbf{v}} = \sup_{\{f_X(x)\}} \frac{J_X^{\mathbf{v}}}{E[C(X)]}$, where $\{f_X(x)\}$ denotes the set of all probability density functions for X such that $E[C(X)] < \infty$. The best worst-case strategy is defined similarly to (4) with $\{f_X(x)\}$ and V replacing $\{p_X(x)\}$ and U, respectively.

The following lemmas are critical in our subsequent analysis.

Lemma 1: For any search strategy $\mathbf{v} \in V$,

$$\sup_{\{f_Y(y)\}} \frac{J_Y^{\mathbf{v}}}{E[C(Y)]} = \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v}}}{C(y)} \ , \tag{5}$$

where $J_Y^{\mathbf{v}}$ is the expected search cost using TTL sequence $\mathbf{v}$ when object location Y has probability density $f_Y(y)$, and $J_y^{\mathbf{v}}$ is the expected search cost using TTL sequence $\mathbf{v}$ when $P(Y = y) = 1$, i.e., a single fixed point.

Proof: We begin by noting that for every $y \in [1, \infty)$, there corresponds a singleton probability density $f_Y(y') = \delta(y' - y)$ so that $P(Y = y) = 1$, and therefore $E[C(Y)] = C(y)$ and $J_Y^{\mathbf{v}} = J_y^{\mathbf{v}}$. We thus have the following inequality

$$\sup_{\{f_Y(y)\}} \frac{J_Y^{\mathbf{v}}}{E[C(Y)]} \geq \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v}}}{C(y)} \ , \tag{6}$$

since the left-hand side is a supremum over a larger set.

On the other hand, setting $A = \sup_{y\in[1,\infty)} \frac{J_y^{\mathbf{v}}}{C(y)}$ we have $\frac{J_y^{\mathbf{v}}}{C(y)} \leq A$ for all $y \in [1, \infty)$. Thus $J_y^{\mathbf{v}} \leq AC(y)$. Then for any random variable Y denoting object location, we can use this inequality along with the independence between $\mathbf{v}$ and Y to obtain:

$$\begin{aligned} \frac{J_Y^{\mathbf{v}}}{E[C(Y)]} &= \frac{\int_{[1,\infty)} J_y^{\mathbf{v}} f_Y(y) \ dy}{\int_{[1,\infty)} C(y) f_Y(y) \ dy} \\ &\leq \frac{\int_{[1,\infty)} AC(y) f_Y(y) \ dy}{\int_{[1,\infty)} C(y) f_Y(y) \ dy} = A \ . \end{aligned} \tag{7}$$

Equation (7) implies that $\frac{J_Y^{\mathbf{v}}}{E[C(Y)]} \le A = \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v}}}{C(y)}$. Since this inequality holds for all possible random variables Y, we have:

$$\sup_{\{f_Y(y)\}} \frac{J_Y^{\mathbf{v}}}{E[C(Y)]} \le \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v}}}{C(y)} \ . \tag{8}$$

Equations (6) and (8) collectively imply the equality in (5), and we have proven Lemma 1. ■

Lemma 2: For any search strategy $\mathbf{u} \in U$,

$$\rho^{\mathbf{u}} = \sup_{\{p_X(x)\}} \frac{J_X^{\mathbf{u}}}{E[C(X)]} = \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} \ , \tag{9}$$

where $J_x^{\mathbf{u}}$ denotes the expected search cost using TTL sequence $\mathbf{u}$ when $Pr(X = x) = 1$, and $\mathbb{Z}^+$ denotes the set of natural numbers and represents all possible singleton object locations.

The proof of this lemma is essentially the same as that of Lemma 1 and is not repeated.

In words, these two lemmas imply that for any TTL sequence, the worst case scenario is when the object location is a constant, i.e., with a singleton probability distribution. We will also subsequently refer to such a single-valued location as a *point*. Note that this constant (i.e., worst case) may not be unique. This result allows us to limit our attention to singleton-valued X and equivalently redefine the minimum worst-case cost ratio ρ^* in equation (4) as $\rho^* = \inf_{\mathbf{u} \in U} \rho^{\mathbf{u}} = \inf_{\mathbf{u} \in U} \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)}$, and similarly for the continuous strategies.

It has been shown in [6] that under the linear cost function $C(u) = \alpha u$, as the network size increases the minimum worst-case cost ratio over all deterministic integer-valued sequences is 4, achieved by the California Split search $\bar{\mathbf{u}} = \{2^{i-1} : i \in \mathbb{Z}^+\} = [1, 2, 4, 8, ...]$. For any cost function $C(\cdot) \in \mathbb{C}$, the real-valued California Split strategy can be defined as a sequence $\mathbf{v}$ satisfying $C(v_1) \in [1, 2)$ and $C(v_i) = 2^{i-1} C(v_1)$ for all $i \in \mathbb{Z}^+$. That is, $\mathbf{v}$ is a sequence with costs growing geometrically by a factor of 2. In the next section we derive randomized strategies that are optimal among *all* admissible strategies. Whereas [6] and [5] derive strategies under linear cost functions, our optimal strategy achieves a much smaller worst-case cost ratio, e, for any cost function $C(\cdot) \in \mathbb{C}$.

IV. Optimal Worst-Case Strategies

In this section, we derive asymptotically optimal continuous and discrete strategies in the limit as the network dimension L approaches ∞. Consequently we will consider TTL sequences of infinite length that are admissible as outlined earlier. The asymptotic case is studied as we are particularly interested in the performance of flooding search in a large network. In addition, it is difficult if at all possible to obtain a general strategy that is optimal for all finite-dimension networks because the optimal TTL sequence often depends on the specific value of L. In this sense, an asymptotically optimal strategy may provide much more insight into the intrinsic structure of the problem. We will see that asymptotically optimal TTL sequences can also perform very well in a network of arbitrary finite dimension.

In what follows we will first derive a tight lower bound on the worst-case cost ratio for continuous strategies. We then introduce a particular randomized continuous strategy that achieves the lower bound, therefore proving that this strategy is optimal in the worst-case. We then repeat the process for the discrete case.

A. A Lower Bound on the Worst-Case Cost Ratio

In deriving a tight lower bound on the worst-case cost ratio, we use Yao's minimax principle [8] and Lemma 2 to obtain the following inequality.

Lemma 3: We have

$$\sup_{\{p_X(x)\}} \inf_{\mathbf{u} \in U^d} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \le \inf_{\mathbf{u} \in U} \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} \ . \tag{10}$$

Proof: For any given object location distribution, the optimal strategy is deterministic. Hence we have $\sup_{\{p_X(x)\}} \inf_{\mathbf{u} \in U^d} \frac{J_X^{\mathbf{u}}}{E[C(X)]} = \sup_{\{p_X(x)\}} \inf_{\mathbf{u} \in U} \frac{J_X^{\mathbf{u}}}{E[C(X)]}$. We also have the following in interchanging the supremum and infimum, see for example [9]:

$$\sup_{\{p_X(x)\}} \inf_{\mathbf{u} \in U} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \le \inf_{\mathbf{u} \in U} \sup_{\{p_X(x)\}} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \ .$$

Combining the above equality and inequality, and using Lemma 2 establishes (10). ■

The corresponding continuous version of Lemma 3 is straightforward with a similar proof.

Lemma 4:

$$\sup_{\{f_X(x)\}} \inf_{\mathbf{v} \in V^d} \frac{J_X^{\mathbf{v}}}{E[C(X)]} \le \inf_{\mathbf{v} \in V} \sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} \ . \tag{11}$$

We now use the above results to first derive a lower bound on the minimum worst-case cost ratio under continuous strategies. Using (11), we note that any lower bound can be found by first selecting a location distribution $f_X(x)$ and deriving the optimal deterministic strategy that minimizes the cost ratio under this distribution. We will assume that the cost function $C(\cdot) \in \mathbb{C}$.

Consider an object location distribution[2] given by $\bar{F}(x) = Pr(X > x) = \left(\frac{C(x)}{C(1)}\right)^{-\alpha}$ for all $x \ge 1$ and some constant $\alpha > 1$. For any deterministic TTL sequence $\mathbf{v} = [v_1, v_2, ...]$, the corresponding expected search cost is given by the following expression, where $v_0 = 1$ is assumed for simplicity of notation:

$$J_X^{\mathbf{v}} = \sum_{j=1}^{\infty} C(v_j) \bar{F}(v_{j-1}) = \sum_{j=1}^{\infty} C(v_j) \left(\frac{C(v_{j-1})}{C(1)}\right)^{-\alpha} \ .$$

Therefore the optimal strategy must satisfy the following partial differential equation:

$$\frac{\partial J_X^{\mathbf{v}}}{\partial v_j} = [C(v_{j-1})^{-\alpha} - \alpha C(v_{j+1}) C(v_j)^{-\alpha-1}] \frac{\partial C(v_j)}{\partial v_j} (C(1))^{\alpha} = 0 \ , \tag{12}$$

[2] A special case of this distribution where cost $C(\cdot)$ is linear, also known as the Zipf distribution, was studied in [6] for which the optimal deterministic strategy was computed. Here we generalize the method to any cost function in $\mathbb{C}$ to derive the class of optimal strategies.

for all $j \geq 1$. Since both the derivative of the cost function and $C(1)$ are strictly positive, for a given fixed v_1 the optimal strategy is to recursively choose v_j that satisfies the following equation for all $j \geq 1$:

$$C(v_{j+1}) = \frac{C(v_j)}{\alpha}\left(\frac{C(v_j)}{C(v_{j-1})}\right)^{\alpha} \ . \tag{13}$$

Note that this optimal sequence satisfies the following:

$$\begin{aligned}\bar{F}(v_j)C(v_{j+1}) &= \left(\frac{C(v_j)}{C(v_{j-1})}\right)^{\alpha}\frac{C(v_j)C(1)^{\alpha}}{\alpha C(v_j)^{\alpha}} \\ &= \frac{C(1)^{\alpha}}{C(v_{j-1})^{\alpha}}\frac{C(v_j)}{\alpha} = \bar{F}(v_{j-1})\frac{C(v_j)}{\alpha} \ . \end{aligned} \tag{14}$$

Summing both sides of (14) from $j = 1$ to $j = \infty$ and multiplying by α gives:

$$\begin{aligned}&\alpha\sum_{j=1}^{\infty}\bar{F}(v_j)C(v_{j+1}) = \sum_{j=1}^{\infty}\bar{F}(v_{j-1})C(v_j) \\ \Longrightarrow\ &\alpha\left(\sum_{i=0}^{\infty}\bar{F}(v_j)C(v_{j+1}) - C(v_1)\right) = \sum_{j=1}^{\infty}\bar{F}(v_{j-1})C(v_j) \ .\end{aligned}$$

Substituting this in the definition of $J_X^{\mathbf{v}}$ gives: $\alpha J_X^{\mathbf{v}} - \alpha C(v_1) = J_X^{\mathbf{v}}$, and thus $J_X^{\mathbf{v}}\frac{\alpha-1}{\alpha} = C(v_1)$. On the other hand, the mean of the object location is as follows, noting that X takes values on $[1,\infty)$:

$$\begin{aligned}E\left[C(X)\right] &= \int_0^{\infty} Pr\left(C\left(X\right) > x\right)\ dx \\ &= C(1) + \int_{C(1)}^{\infty}\bar{F}(C^{-1}(x))\ dx \\ &= C(1) + \int_{C(1)}^{\infty}\left[\frac{C(C^{-1}(x))}{C(1)}\right]^{-\alpha}\ dx = \frac{\alpha}{\alpha-1}C(1) \ .\end{aligned}$$

The above imply that for a sequence defined by a given v_1 and following recursion (13), the cost ratio is

$$\frac{J_X^{\mathbf{v}}}{E\left[C\left(X\right)\right]} = J_X^{\mathbf{v}}\frac{(\alpha-1)}{\alpha C(1)} = \frac{C(v_1)}{C(1)}. \tag{15}$$

This means that for a given α, the sequence that generates the smallest cost ratio will follow the recursion (13) and use the smallest possible value of v_1. However, not all values of v_1 lead to an increasing sequence $\mathbf{v}$. In fact, we have the following result:

Lemma 5: Consider any infinite length sequence $\mathbf{v} = [v_1, v_2, ..]$, where v_1 is some positive constant and v_k, $k \geq 2$, is generated by the recursion given by (13). Then $\mathbf{v}$ is an increasing sequence if and only if the following condition holds:

$$\frac{C(v_1)}{C(1)} \geq \alpha^{\left(\sum_{k=1}^{\infty}\alpha^{-k}\right)} = \alpha^{\frac{1}{\alpha-1}}. \tag{16}$$

The proof of this lemma can be found in the Appendix.

Therefore we can achieve a minimum cost ratio value of $\alpha^{\frac{1}{\alpha-1}}$ by using a TTL sequence defined by recursion (13) and v_1 such that $\frac{C(v_1)}{C(1)}$ is $\alpha^{\frac{1}{\alpha-1}}$. When $\alpha > 1$, $\alpha^{\frac{1}{\alpha-1}}$ is a decreasing function of α, with its maximum achieved as α approaches 1 from above. In addition we have $\lim_{\alpha\to 1^+}\alpha^{\frac{1}{\alpha-1}} = e$, which follows from the definition of the exponential constant.

Therefore using (11) we have obtained a lower bound on the worst-case cost ratio, given by the next lemma.

Lemma 6: For any $C(\cdot) \in \mathbb{C}$, the worst-case cost ratio of any continuous strategy is lower-bounded by e, i.e.:

$$\inf_{\mathbf{v}\in V}\sup_{x\in[1,\infty)}\frac{J_x^{\mathbf{v}}}{C(x)} \geq e \ . \tag{17}$$

This result implies that if we can obtain a TTL sequence whose worst-case ratio is e, then it must be an optimal worst-case strategy. We derive such a strategy in the next two subsections.

B. A Class of Jointly Defined Randomized Strategies

Definition 2: Assume that the cost function $C(\cdot) \in \mathbb{C}$. Let $\mathbf{v}[r, F_{v_1}(x)]$ denote a jointly defined sequence $\mathbf{v} = [v_1, v_2, ...]$ with a configurable parameter r, generated as follows:

(J.1) The first TTL value v_1 is a continuous random variable taking values in the interval $\left[1, C^{-1}(rC(1))\right)$, with its cdf given by some nondecreasing, right-continuous function $F_{v_1}(x) = Pr(v_1 \leq x)$. Note that this means $F_{v_1}(1) = 0$ and $F_{v_1}\left(C^{-1}(rC(1))\right) = 1$.

(J.2) The k-th TTL value v_k is defined by $v_k = C^{-1}\left(r^{k-1}C(v_1)\right)$ for all positive integers k.

From (J.1) and (J.2), it can be seen that r and $F_{v_1}(x)$ uniquely define the TTL strategy. Note that successive $C(v_k)$ form a geometric sequence with power r.

Lemma 7: Consider any strategy $\mathbf{v}[r, F_{v_1}(x)]$ constructed using steps $(J.1)$ and (J.2) above. Assume $C(\cdot) \in \mathbb{C}$. Let $\bar{F}_{v_1}(y) = 1 - F_{v_1}(y)$. Then the worst-case cost ratio is given by:

$$\sup_{x\in[1,\infty)}\frac{J_x^{\mathbf{v}}}{C(x)} = \sup_{1\leq z<r}\left\{\frac{r}{r-1}\frac{h(r)+(r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)}\right\}$$

where $h'(z)$ denotes the derivative of h with respect to z, and $h(z)$ is defined as follows for $1 \leq z < r$:

$$h(z) = C(1) + \int_{C(1)}^{zC(1)}\bar{F}_{v_1}(C^{-1}(y))\ dy \ . \tag{18}$$

The proof is given in the appendix. This lemma reduces the space over which the supremum is taken in order to calculate the worst-case cost ratio.

C. An Optimal Continuous Strategy

For $1 \leq z \leq r$ and a given strategy $\mathbf{v}[r, F_{v_1}(x)]$, define $\Phi(z)$ as follows:

$$\Phi(z) = \frac{r}{r-1}\frac{h(r)+(r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} \ . \tag{19}$$

From Lemma 7, the worst-case cost ratio of $\mathbf{v}$ is the supremum of $\Phi(z)$ over the range $1 \leq z < r$. The following four boundary conditions are true for any function $h(z)$ as defined by (18): $h(1) = C(1)$, $h(r) = E[C(v_1)]$, $h'(1) = C(1)$, and $h'(r) = 0$.

Theorem 1: Assume $C(\cdot) \in \mathbb{C}$. We have

$$\inf_{\mathbf{v}\in V}\sup_{x\in[1,\infty)}\frac{J_x^{\mathbf{v}}}{C(x)} = e \ . \tag{20}$$

Moreover, this worst-case cost ratio is obtained by strategy $\mathbf{v}^*[e, \ln\frac{C(x)}{C(1)}]$. In other words, the optimal strategy is defined as follows: v_1^* has the cdf $F_{v_1^*}(x) = \ln \frac{C(x)}{C(1)}$ for $1 \le x < C^{-1}(eC(1))$, and $v_k^* = C^{-1}\left(e^{k-1}C\left(v_1\right)\right)$ for all positive integers k.

Proof: Consider strategy $\mathbf{v}^*$ as described. Note that because $F_{v_1^*}(x) = \ln \frac{C(x)}{C(1)}$ and $r = e$, we have:

$$\begin{aligned} h(z) &= C(1) + \int_{C(1)}^{zC(1)} \bar{F}_{v_1^*}\left(C^{-1}(y)\right)\, dy \\ &= C(1) + \int_{C(1)}^{zC(1)} \left(1 - \ln \frac{y}{C(1)}\right)\, dy \\ &= C(1)\left[z - z(\ln z - 1) - 1\right] = C(1)\left[2z - z\ln z - 1\right] \ . \end{aligned}$$

Note that $h(e) = C(1)\,(e-1)$ and $h'(z) = C(1)\,[1 - \ln z]$. Thus we have for $1 \le z < r$:

$$\Phi(z) = \frac{e}{e-1} \frac{C(1)(e-1) + (e-1)C(1)\left[2z - z\ln z - 1\right]}{zC(1)} - \frac{eC(1)(1-\ln z)}{C(1)} = e \ .$$

Hence it is clear from Lemma 7 that the worst-case cost ratio of this sequence is e. Combine this with Lemma 6 which showed the worst-case cost ratio of any continuous strategy is lower bounded by e, we complete the proof. ■

As an example, when the cost is linear, i.e. $C(x) = x$ for all x, the optimal strategy $\mathbf{v}^* = [v_1^*, v_2^*, ...]$ is as follows. The first TTL value is a random variable v_1^* with cdf $F_{v_1^*}(z) = \ln z$ for $1 \le z < e$. Successive TTL values are defined as $v_k^* = e^{k-1} v_1^*$.

The above optimal strategy belongs to the family of strategies given by $\mathbf{v}[r, \frac{1}{\ln r} \ln \frac{C(x)}{C(1)}]$, indexed by the parameter r, with $r = e$ being the optimal strategy. There is an interesting interpretation of this family of strategies. Specifically, if Z denotes a random variable uniformly distributed in the interval $[0, 1)$, then this strategy has costs satisfying $C(v_k) = C(1) r^{k-1} r^Z$ for all $k \ge 1$. This interpretation is used in Section V.

Note that the minimum cost-ratio derived in Theorem 1 is the same for all cost functions in $\mathbb{C}$. The reason for this will become clearer in Section VI when we show an equivalence result between different cost functions.

D. An Optimal Discrete Strategy

For the discrete case, the minimum worst-case cost appears to have a stronger dependence on the specific cost function. We therefore limit attention to the following subclass of $\mathbb{C}$.

Definition 3: A function $C(\cdot) \in \mathbb{C}$ belongs to the class $\mathbb{C}_q$ for some $q \ge 1$ if $\lim_{x\to\infty} \frac{C(x+1)}{C(x)} = q$ and $\frac{C(x+1)}{C(x)} \ge q$ for all $x \in [1, \infty)$.

Note that $\mathbb{C}_1$ contains all polynomial cost functions. The case of $q > 1$ includes for example exponential cost functions of the form q^x. Therefore this definition remains quite general even though it is a subclass of $\mathbb{C}$. We first derive a lower-bound on the best worst-case cost ratio, by utilizing the next lemma. Let X_α denote the random variable with tail distribution $Pr(X_\alpha > x) = [C(x)/C(1)]^{-\alpha}$ for all $x \ge 1$ and some $\alpha > 1$. Then:

Lemma 8: If $C(\cdot) \in \mathbb{C}_q$, then $\lim_{\alpha\to 1^+} \frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} = q$.

Proof: Because $C(x+1) \ge qC(x)$ for all x, we have $C(X_\alpha + 1) \ge qC(X_\alpha)$ *w.p.* 1. Hence, $\frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} \ge q$ for all α. Hence to complete the proof, we need to show that for any $\epsilon > 0$, there exists $\bar{\alpha}$ such that $\frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} < q + \epsilon$ for all $1 < \alpha < \bar{\alpha}$. To begin, fix $\epsilon > 0$. Since $C(\cdot) \in \mathbb{C}_q$, there exists x^* such that $\frac{C(x+1)}{C(x)} < q + \frac{\epsilon}{2}$ for all $x > x^*$. Let $I(\cdot)$ denote the indicator function such that $I(A) = 1$ if A is true and 0 otherwise. We have:

$$\begin{aligned} E[C(X_\alpha+1)I(X_\alpha > x^*)] &< \left(q + \frac{\epsilon}{2}\right) E[C(X_\alpha)I(X_\alpha > x^*)] \\ &\le \left(q + \frac{\epsilon}{2}\right) E[C(X_\alpha)] \end{aligned} \tag{21}$$

At the same time we also have:

$$\lim_{\alpha\to 1^+} \frac{E[C(X_\alpha+1)I(X_\alpha \le x^*)]}{E[C(X_\alpha)]} \le \lim_{\alpha\to 1^+} \frac{C(x^*+1)}{E[C(X_\alpha)]} = 0 \ ,$$

since $C(x^*+1) < \infty$ and $E[C(X_\alpha)] = \frac{\alpha}{\alpha-1}$. Hence, there exists an $\bar{\alpha}$ such that for all $1 < \alpha < \bar{\alpha}$ we have $E[C(X_\alpha+1)I(X_\alpha \le x^*)] < \frac{\epsilon}{2} \cdot E[C(X_\alpha)]$. Combining this with (21) gives for $1 < \alpha < \bar{\alpha}$:

$$\begin{aligned} &E[C(X_\alpha+1)] \\ &\quad = E[C(X_\alpha+1)I(X_\alpha > x^*)] + E[C(X_\alpha+1)I(X_\alpha \le x^*)] \\ &\quad < \left(q + \frac{\epsilon}{2} + \frac{\epsilon}{2}\right) E[C(X_\alpha)] = (q+\epsilon)E[C(X_\alpha)] \ , \end{aligned}$$

which completes the proof. ■

This result can be used to obtain the following lemma.

Lemma 9: For $C(\cdot) \in \mathbb{C}_q$, the best worst-case cost ratio is lower-bounded by $\inf_{\mathbf{u}\in U} \sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} \ge \frac{e}{q}$.

Proof: Fix some $\mathbf{u} \in U$. For any integer $x \ge 2$, $\frac{J_x^{\mathbf{u}}}{C(x)} = \lim_{\epsilon\to 0} \frac{J_{x-1+\epsilon}^{\mathbf{u}}}{C(x+\epsilon)} = \sup_{y\in[x-1,x)} \frac{J_y^{\mathbf{u}}}{C(y+1)}$, because $J_{x-1+\epsilon}^{\mathbf{u}} = J_x^{\mathbf{u}}$ for all $0 < \epsilon \le 1$, and $C(\cdot)$ is increasing. Hence:

$$\sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} = \sup\left\{ \frac{J_1^{\mathbf{u}}}{C(1)}, \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{u}}}{C(x+1)} \right\} \tag{22}$$

In order to find a lower-bound to the above worst-case ratio, we first examine all strategies $\mathbf{v} \in V$. It can be shown, similarly to Lemma 1 that $\sup_{\{f_X(X)\}} \frac{J_X^{\mathbf{v}}}{E[C(X+1)]} = \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x+1)}$. Thus similarly to Lemma 4, we have:

$$\sup_{\{f_X(x)\}} \inf_{\mathbf{v}\in V^d} \frac{J_X^{\mathbf{v}}}{E[C(X+1)]} \le \inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x+1)} \ . \tag{23}$$

Any lower bound can be found by first selecting a distribution $f_X(x)$ and deriving the optimal fixed strategy.

Consider the random variable X_α as defined earlier for some fixed $\alpha > 1$. It was shown in Section IV-A that for object location X_α, the optimal cost ratio is $\alpha^{\frac{1}{\alpha-1}}$, which approaches e as $\alpha \to 1^+$. Hence using Lemma 8 we have:

$$\begin{aligned} &\lim_{\alpha\to 1^+} \inf_{\mathbf{v}\in V^d} \frac{J_{X_\alpha}^{\mathbf{v}}}{E[C(X_\alpha+1)]} \\ &= \lim_{\alpha\to 1^+} \frac{E[C(X_\alpha)]}{E[C(X_\alpha+1)]} \inf_{\mathbf{v}\in V^d} \frac{J_{X_\alpha}^{\mathbf{v}}}{E[C(X_\alpha)]} = \frac{e}{q} \ . \end{aligned}$$

Using this result in (23) and $U \subset V$ gives us:

$$\inf_{\mathbf{u}\in U} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{u}}}{C(x+1)} \ge \inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x+1)} \ge \frac{e}{q}$$

From (22), the left-hand side is less than or equal to $\inf_{\mathbf{u}\in U}\sup_{x\in\mathbb{Z}^+}\frac{J_x^{\mathbf{u}}}{C(x)}$, thus completing the proof. ■

This result says that if we can find a discrete strategy whose worst-case cost ratio is e/q for $C(\cdot)\in\mathbb{C}_q$, then this strategy must be optimal among all strategies in U. Unfortunately it appears difficult to find strategies matching this lower bound for all $C(\cdot)\in\mathbb{C}_q$. The reason appears to be that for large q, the cost function value grows very rapidly and thus it becomes harder to find strategies that match this bound. It is, however, possible to do so for the special case of $q=1$, as shown in the next theorem.

Theorem 2: For $C(\cdot)\in\mathbb{C}_1$, we have:

$$\inf_{\mathbf{u}\in U}\sup_{x\in\mathbb{Z}^+}\frac{J_x^{\mathbf{u}}}{C(x)}=e\ .$$

Moreover, this worst-case cost ratio is obtainable by strategy $\mathbf{u}^*$, which is constructed as follows. Consider the continuous strategy $\mathbf{v}^*[e,\ln\frac{C(x)}{C(1)}]$, and set $u_k^*=\lfloor v_k^*\rfloor$ for all k to obtain the discrete strategy $\mathbf{u}^*=[u_1^*,u_2^*,...]$.

Proof: Consider strategies $\mathbf{u}^*$ and $\mathbf{v}^*$ as described in the theorem. Lemma 9 implies $\sup_{x\in\mathbb{Z}^+}\frac{J_x^{\mathbf{u}^*}}{C(x)}\geq e$. Thus, to complete the proof we need to show $\sup_{x\in\mathbb{Z}^+}\frac{J_x^{\mathbf{u}^*}}{C(x)}\leq e$.

Fix any $x\in\mathbb{Z}^+$. Note that $x>\lfloor v_k^*\rfloor$ if and only if $x>v_k^*$, since x is an integer. Therefore for all k, $I(x>\lfloor v_k^*\rfloor)=I(x>v_k^*)$ $w.p.$ 1. In addition, $C(\cdot)$ being increasing implies $C(\lfloor v_k^*\rfloor)\leq C(v_k^*)$ $w.p.$1. Therefore,

$$\begin{aligned}J_x^{\mathbf{u}^*}&=E\left[\sum_{k=1}^{\infty}I(x>\lfloor v_{k-1}^*\rfloor)C(\lfloor v_k^*\rfloor)\right]\\&\leq E\left[\sum_{k=1}^{\infty}I(x>v_{k-1}^*)C(v_k^*)\right]=J_x^{\mathbf{v}^*}\leq eC(x)\quad,\end{aligned}$$

where the last inequality holds because the worst-case cost ratio for $\mathbf{v}^*$ is e as proven in Theorem 1. Since this result holds for all integers x, we have $\sup_{x\in\mathbb{Z}^+}\frac{J_x^{\mathbf{u}^*}}{C(x)}\leq e$. ■

Since $\mathbb{C}_1$ includes all increasing polynomials, the optimal strategy given in Theorem 2 can be used when the cost is given by or can be approximated by a polynomial function, which is not a very restrictive assumption.

V. Performance Comparison and Discussion

In this section we first compare the performance of the optimal randomized strategy with the optimal non-random strategy and illustrate the fundamental reason behind why randomized strategies result in lower worst-case cost ratio. We then consider other performance measures for evaluating randomized search strategies.

A. A Comparison between Randomized and Deterministic Strategies

In Figure 1 we compare the cost ratio of the optimal discrete strategy given by Theorem 2 to that of the non-random TTL sequence given by the California Split search $u_k=2^{k-1}$ for all k under the linear cost function $C(k)=k$. We see that the cost ratio oscillates for the fixed TTL sequence while randomization essentially has the *averaging* effect that "smooths out" the cost ratio across neighboring locations/points. In fact the curve of the optimal continuous strategy does not have local minima or maxima. One may view this as the built-in *robustness* of a randomized policy for the underlying criterion of worst-case performance. Also note that the worst-case cost ratio e is reached asymptotically from below as $L\to\infty$, and hence the cost ratio at any finite object location is less than the worst-case cost ratio.

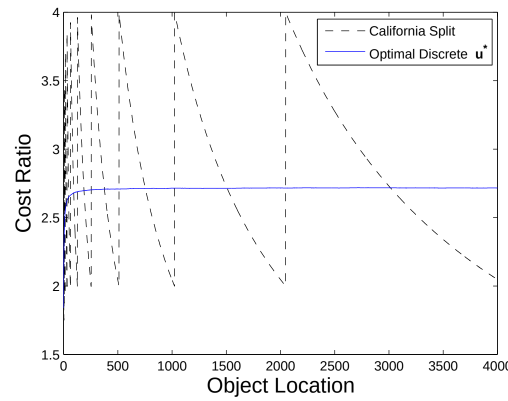


Fig. 1. Cost ratio as a function of object location for the optimal discrete sequence $\mathbf{u}^*$ described in Theorem 2, and California Split search defined by $u_k=2^{k-1}$ for all k. Cost is assumed to be linear.

The optimal randomized strategy $\mathbf{v}^*[e,\ln\frac{C(x)}{C(1)}]$ is essentially the best among the following family of strategies that achieve a similar flat cost ratio curve as shown above, given by $\mathbf{v}[r,\frac{1}{\ln r}\ln\frac{C(x)}{C(1)}]$. It is not difficult to show that the worst-case cost ratio of this family of strategies is $\frac{r}{\ln r}$, which occurs as the object location goes to infinity. By differentiating and noting convexity, $r=e$ minimizes the worst-case cost ratio, and the minimum is e.

B. Other Performance Measures

Next we discuss alternative performance measures for analyzing randomized search strategies. We will again assume that $C(\cdot)\in\mathbb{C}$, and begin with continuous strategies.

The performance measure we have been using is the worst-case cost ratio with respect to an oblivious adversary, who knows the strategy but not the realization of the strategy. As pointed out in Section III, the lower bound e on the worst-case cost ratio does not necessarily bound the cost ratio for all realizations of X and strategy $\mathbf{v}$. This leads us to consider the competitive ratio with respect to an *adaptive offline adversary* [8] who knows the *realization* of the real-valued strategy $\mathbf{v}$ for every search. Let the *worst-realization cost ratio* $\Gamma_X^{\mathbf{v}}$ denote the maximum (over all realizations of strategy $\mathbf{v}$) cost ratio for strategy $\mathbf{v}$ when the object location is a random variable X. Specifically, $\Gamma_X^{\mathbf{v}}=\sup_{\tilde{\mathbf{v}}\in\Upsilon^{\mathbf{v}}}\frac{J_X^{\tilde{\mathbf{v}}}}{E[C(X)]}$, where $\Upsilon^{\mathbf{v}}$ denotes the set of all possible realizations of strategy $\mathbf{v}$. Then the performance of a search strategy against an adaptive offline adversary can be measured by the following competitive ratio, which is the *worst-case worst-realization cost ratio*, denoted by $\Gamma^{\mathbf{v}}$:

$$\Gamma^{\mathbf{v}} = \sup_{\{f_X(x)\}} \Gamma^{\mathbf{v}}_X = \sup_{x\in[1,\infty)} \Gamma^{\mathbf{v}}_x \ . \tag{24}$$

The second equality can be shown in a manner similar to the proof of Lemma 1. To distinguish, we will refer to $\rho^{\mathbf{v}}$ discussed in the previous sections as the *worst-case average cost ratio*.

As discussed in [8], the minimum obtainable competitive ratio with respect to an adaptive offline adversary is the same as the minimum worst-case average cost ratio of all deterministic strategies. The latter under $C(\cdot) \in \mathbb{C}$ can be shown to be 4, same as in the linear cost function case [3]. Therefore, we have that $\inf_{\mathbf{v}\in V} \Gamma^{\mathbf{v}} = 4$.

Similarly, let $\gamma^{\mathbf{v}}_X$ and $\gamma^{\mathbf{v}}_x$ denote the *best-realization cost ratio* of strategy $\mathbf{v}$ when object location is a random variable X or a single point $x \in [1,\infty)$, respectively. These definitions for best and worst realizations are easily extendable to integer-valued strategies $\mathbf{u} \in U$ by replacing the possible set of locations $[1,\infty)$ with $\mathbb{Z}^+$. Finally, let $\Lambda^{\mathbf{v}}_x$ denote the variance of the search cost of strategy $\mathbf{v}$ with fixed object location $x \in [1,\infty)$. Therefore, $\Lambda^{\mathbf{v}}_x/C(x)^2$ (which we refer to as the *cost ratio variance*) is the variance of the ratio $j^{\mathbf{v}}_x/C(x)$ when object location is x.

In the following proposition, we list these quantities for the class of jointly defined continuous strategies $\mathbf{v}[r, F_{v_1}(x)]$ given by Definition 2, and in particular when $F_{v_1}(x) = \frac{1}{\ln r}\ln\frac{C(x)}{C(1)}$.

Proposition 1: For any real-valued randomized strategy $\mathbf{v}[r, F_{v_1}(x)]$ given by Definition 2, $\Gamma^{\mathbf{v}} \le \frac{r^2}{r-1}$. When $F_{v_1}(x) = \frac{1}{\ln r}\ln\frac{C(x)}{C(1)}$, we have $\Gamma^{\mathbf{v}} = \frac{r^2}{r-1}$. Under this distribution, the best-case realization cost ratio is given by: $\gamma^{\mathbf{v}}_x = \frac{r-r^{-k+1}}{r-1}$. Finally, the asymptotic cost ratio variance of the strategy under this distribution is given by: $\lim_{x\to\infty}\frac{\Lambda^{\mathbf{v}}_x}{C(x)^2} = \frac{r^4-r^2}{2(\ln r)(r-1)^2} - \frac{r^2}{(\ln r)^2}$.

These calculations are derived in the appendix. The cost ratio, under $C(\cdot) \in \mathbb{C}$, of the optimal continuous strategy is depicted in Figure 2 (TOP) as a function of object location cost.

Performance of strategies of the type $\mathbf{v}[r, F_{v_1}(x)]$ where $F_{v_1}(x) = \frac{1}{\ln r}\ln\frac{C(x)}{C(1)}$ are shown in Figure 2 (BOTTOM) as functions of r. As can be seen, we can appropriately select the value of r depending on whether the goal is to minimize the worst-case expected cost or the worst-case worst-realization cost. As we have mentioned earlier, an interpretation of this family of random strategies is that their costs satisfy $C(v_k) = C(1)r^{k-1}r^Z$, where Z is a random variable uniformly distributed in interval $[0,1)$. Thus for fixed r, every realization of $\mathbf{v}$ is a nonrandom sequence with costs increasing geometrically by a factor of r. Therefore, the asymptotic worst and best realization cost ratios of $\mathbf{v}$ match that of the corresponding deterministic geometric strategies. In particular, we note that by using $r = 2$, we can obtain a worst-case worst-realization cost ratio of 4. This is precisely because any realization of this strategy is simply the deterministic California Split search[4]. At the same time, this random strategy obtains a worst-case average cost ratio of approximately 2.8854. Therefore this particular strategy strictly outperforms the deterministic California Split search.

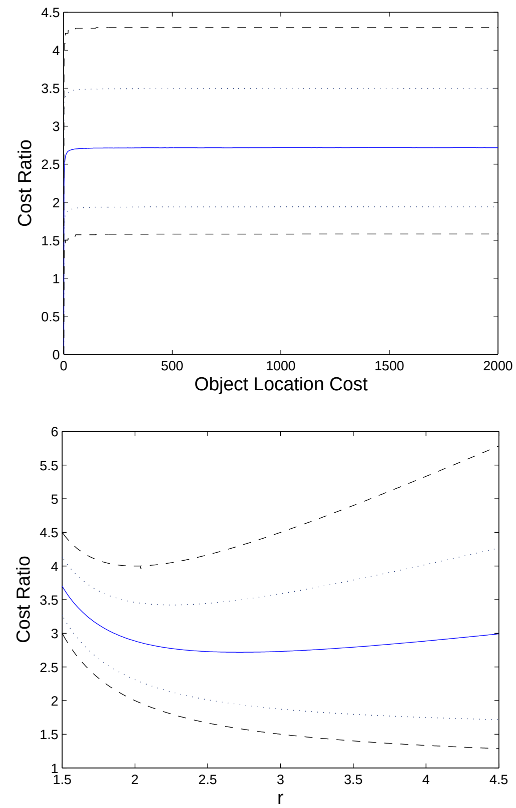


Fig. 2. (TOP): Performance of the optimal continuous strategy (Theorem 1) as a function of object location cost. Worst and best realization cost ratio (top and bottom dashed lines), average cost ratio (solid), and average cost ratio $+/-$ one standard deviation (dotted) are shown. (BOTTOM): Performance of $\mathbf{v}\left[r, \frac{1}{\ln r}\ln\frac{C(x)}{C(1)}\right]$ as a function of r. Worst-case average cost ratio (solid), asymptotic worst and best realization cost ratio (dashed), and worst-case average cost $+/-$ standard deviation (dotted) are shown.

Similar analysis can be carried out for discrete strategies. The calculations are rather tedious and do not provide any more insight. The performance of the discrete strategy is very similar to its continuous version with respect to the performance measures discussed in this section and is therefore not shown separately due to space limitations.

C. Comparison with Optimal Average Cost Strategies

The worst-case cost ratio we have been using so far is in general a conservative/pessimistic performance measure. As mentioned earlier, if the probability distribution of the location of the object is known *a priori*, then we can derive the optimal strategy that achieves the lowest average cost for the given object distribution, using a dynamic programming formulation

[3]In Section VI we establish an equivalence relationship between linear and general cost functions, which can be used to show that 4 is also the minimum worst-case cost ratio among deterministic strategies under general cost functions.

[4]We note, however, the worst-realization cost ratio of 4 here does not depend on the random variable Z being uniform. It can be seen that regardless of Z, the costs $C(v_k)$ will be a geometric sequence of factor 2 as long as $r = 2$.

[5]. On the other hand, the optimal average-cost strategy can potentially be highly sensitive to small disturbances to our knowledge about the object location distribution, while worst-case strategies may be more robust.

We compare the two under the following example scenarios. Consider a network of finite dimension L and the linear cost function $C(k) = k$. We examine what happens when there are errors in our estimate of the location distribution. Consider when the object location has probability mass function $P(X = x) = \beta x^{\alpha}$ for all $1 \leq x \leq L$, where the constant α defines the distribution and β is a normalizing constant. Note that $\alpha = 0$ corresponds to uniform location distribution. We let DP(α') denote the optimal (deterministic) average-cost strategy derived using dynamic programming when assuming $\alpha = \alpha'$ in the distribution of X. We then compute the expected search cost of DP(0) and DP(-2.5) when the location distribution is in fact defined by some other α, for $-10 \leq \alpha \leq 10$. Similarly, we calculate the average search cost under these distributions when using the optimal worst-case (randomized) strategy, RAND.

These results are shown in Figure 3. In Figure 3 (TOP), the average cost of DP(0) and RAND strategies are shown as functions of L for $\alpha = -1$, 0, and 1. In Figure 3 (BOTTOM) the performance of these two strategies and DP(-2.5) are plotted for $L = 100$ as functions of α. As can be seen, DP(0) is more robust (less sensitive in the change in α) than RAND, while for DP(-2.5) the opposite is true. For small (negative) α, RAND outperforms DP(0) and in some cases the average-cost of DP(0) is 38 times larger. On the other hand, for large (positive) α, DP(0) is better, but the average-cost of RAND is greater only by a factor of 1.3. Thus we see that the dynamic programming strategy should only be used if we are fairly certain about the object location distribution.

This quantitative relationship obviously varies with the underlying assumptions on the location distribution and the errors introduced. This specific example nonetheless illustrates the general trade-off between search cost and robustness.

D. Potential Limitation

The construction of the optimal continuous and discrete strategies derived in the previous section depends on our ability to define and invert a cost function that is defined for all $x \in [1, \infty)$. While conceptually and fundamentally appealing, this construction may pose a problem in practice. If the search cost is only known for integer TTL values, then in order to obtain the optimal discrete search strategy given in Theorem 2, we would need to interpolate and create an increasing, differentiable, and continuous cost function defined over the positive real line. Such a process is not always easy to carry through.

Motivated by this, in Section VII we discuss strategies that are sub-optimal with respect to our performance measure but still outperform deterministic strategies and that could be easier to derive and implement than those introduced in Section IV. We first establish in the next section an equivalence relationship between the linear cost function and a general cost function. With this result our later discussion is greatly simplified.

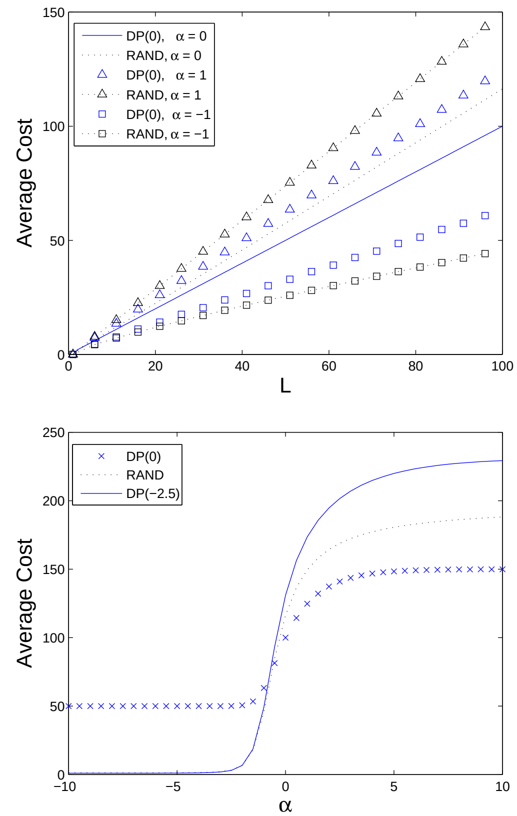


Fig. 3. (TOP): Comparison between DP(0) and RAND for varying L and α. (BOTTOM): Performance of DP(0), DP(-2.5) and RAND as functions of α when $L = 100$.

VI. An Equivalence Result between Linear and General Cost Functions

In this section, we present a mapping that establishes the equivalency between real-valued TTL sequences under different cost functions.

Lemma 10: Let $J_x^{\mathbf{w},l}$ denote the search cost of using strategy $\mathbf{w} = [w_1, w_2, ...]$ when the cost function is linear and object location is x for some $x \in [1, \infty)$. Consider any cost function $C(\cdot) \in \mathbb{C}$. Let $\mathbf{v}$ denote the strategy that is constructed as $\mathbf{v} = C^{-1}(\mathbf{w} \cdot C(1))$, i.e., $v_k = C^{-1}(w_k \cdot C(1))$ for all positive integers k. Let $J_x^{\mathbf{v},g}$ denote the search cost of using strategy $\mathbf{v} = [v_1, v_2, ...]$ when the object location is x for some $x \in [1, \infty)$ and the cost function is $C(x)$. Then we have the following:

$$\sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{w},l}}{x} = \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v},g}}{C(y)} \tag{25}$$

Proof: Fix any $y \in [1, \infty)$. We are given that for all k, $C(v_k) = w_k \cdot C(1)$ $w.p.1$. Combining this with $C(\cdot)$ being positive and strictly increasing, we have $I(y > v_k) = I\left(C(y)/C(1) > w_k\right)$ $w.p.1$ for all k. Therefore, by letting $x = C(y)/C(1)$ and noting that $x \in [1, \infty)$, we have the following:

$$\begin{aligned} J_y^{\mathbf{v},g} &= E\left[\sum_{k=1}^{\infty} I(y > v_{k-1}) C(v_k)\right] \\ &= E\left[\sum_{k=1}^{\infty} I\left(\frac{C(y)}{C(1)} > w_{k-1}\right) w_k \cdot C(1)\right] \\ &= C(1)\cdot E\left[\sum_{k=1}^{\infty} I(x > w_{k-1}) w_k\right] = C(1) J_x^{\mathbf{w},l} \end{aligned}$$

Thus we have that for all $y \in [1,\infty)$, $\exists\ x \in [1,\infty)$ such that $\frac{J_y^{\mathbf{v},g}}{C(y)} = \frac{C(1)J_x^{\mathbf{w},l}}{C(y)} = \frac{J_x^{\mathbf{w},l}}{x}$. Since this result holds for all $y \in [1,\infty)$, (25) follows. ■

Lemma 10 implies that for any strategy $\mathbf{w}$ under linear cost there corresponds a strategy $\mathbf{v}$ that has the same performance under cost functions $C(\cdot) \in \mathbb{C}$. Hence this result helps to explain why the minimum worst-case cost ratio derived in Theorem 1 is the same for all cost functions in this general class. Note that this mapping in its precise form only applies to continuous strategies. For discrete strategies, approximations can be made to obtain similar strategies as described in the next section.

As an application of this mapping, consider a continuous strategy $\mathbf{w}$ (under linear cost) in which the TTL random variables are continuous and the k-th TTL value has probability density function $f_{w_k}(x)$ defined for all $x \in [1,\infty)$. From Lemma 10, the strategy $\mathbf{v} = C^{-1}(\mathbf{w}\cdot C(1))$ has the same worst-case cost ratio under cost function $C(\cdot) \in \mathbb{C}$. The k-th TTL random variable v_k therefore has probability density function f_{v_k} defined as follows for all $y \in [1,\infty)$:

$$f_{v_k}(y) = f_{w_k}\left(\frac{C(y)}{C(1)}\right) \cdot \frac{dC(y)}{dy}\frac{1}{C(1)}\ . \tag{26}$$

When v_k's are mutually independent, (26) for all k uniquely defines the strategy $\mathbf{v}$.

VII. UNIFORM RANDOMIZATION

In [5] we introduced a class of *uniformly randomized strategies*. We derived the optimal strategy within this class under the linear cost function assumption, and showed that the best worst-case cost ratio is 2.9142. Below we briefly summarize these results and extend them to more general cost functions.

A. Results on Uniform Randomization

Definition 4: For any infinite, increasing sequence $\mathbf{b} = [b_1, b_2, ...]$ in which the elements b_k are positive integers and $b_j > b_k$ for all $j > k$, a *uniformly randomized* TTL sequence $\mathbf{u} = [u_1, u_2, ...]$ is created by assigning the following probability distribution to each TTL random variable u_k:

$$Pr(u_k = l) = \begin{cases} \frac{1}{b_{k+1}-b_k} & \text{if} \quad b_k \le l \le b_{k+1}-1 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where l is any positive integer.

Essentially the elements in the nonrandom sequence $\mathbf{b} = [b_1, b_2, ...]$ serve as the boundaries of a sequence of non-overlapping ranges over which each random variable u_k is uniformly distributed. These ranges collectively cover all positive integers. Following this definition, for each nonrandom TTL sequence, there exists a corresponding uniformly randomized version.

Theorem 3: (From [5]) Let U' denote the set of all non-random and uniformly randomized TTL sequences. Then:

$$\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}} = \inf_{\mathbf{u}\in U'} \sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{x} = \frac{3}{2} + \sqrt{2} \approx 2.9142\ . \tag{28}$$

Furthermore, this ratio is achieved by the uniformly randomized strategy defined by the boundary sequence $b_k = \lfloor r^{k-1} \rfloor$ with $r = \sqrt{2}+1$.

We have analogous results when extending the set of admissible strategies to V.

Definition 5: For any infinite, increasing fixed sequence $\mathbf{b} = [b_1, b_2, ...]$ in which the elements b_k are positive real numbers and $b_j > b_k \ge 1$ for all $j > k$, a *uniformly randomized* continuous-valued TTL sequence $\mathbf{v} = [v_1, v_2, ...]$ is created by assigning the following probability density f_{v_k} to each TTL random variable v_k:

$$f_{v_k}(y) = \begin{cases} \frac{1}{b_{k+1}-b_k} & \text{if} \quad b_k \le y < b_{k+1} \\ 0 & \text{otherwise} \end{cases}\ . \tag{29}$$

It can be shown that for such uniformly randomized continuous-valued TTL sequences, we have the following result which is similar to Theorem 3.

Theorem 4: Let V' denote the set of all nonrandom and uniformly randomized continuous-valued TTL sequences. Then:

$$\inf_{\mathbf{v}\in V'} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{x} = \frac{3}{2} + \sqrt{2} \approx 2.9142\ . \tag{30}$$

Furthermore, this ratio is achieved by the uniformly randomized continuous strategy defined by the boundary sequence $b_k = r^{k-1}$ where $r = \sqrt{2}+1$.

The proof of Theorem 4 is very similar to that of Theorem 3. The latter can be found in [5].

Using Lemma 10 we can obtain a discrete strategy $\mathbf{u}$ which performs similarly (under any increasing cost function) as the optimal uniformly randomized sequence under the linear cost function. We first show an example when the cost is quadratic, i.e., $C(x) = \alpha x^2$. To begin, consider the optimal continuous uniformly randomized TTL strategy $\mathbf{w}$ with boundary values given by $b_k^{\mathbf{w}} = r^{k-1}$ with $r = \sqrt{2}+1$, and construct a uniformly randomized strategy $\hat{\mathbf{w}}$ with boundary values $b_k^{\hat{\mathbf{w}}} = \lfloor r^{\frac{k-1}{2}} \rfloor^2$. To create the corresponding strategy $\mathbf{v}$ under the quadratic cost function, we use equation (26) to determine the probability distribution of each TTL random variable. In particular:

$$f_{v_k}(y) = \frac{2y}{b_{k+1}^2 - b_k^2} \quad \text{if } b_k \le y < b_{k+1}\ , \tag{31}$$

where $b_k = \sqrt{b_k^{\hat{\mathbf{w}}}} = \lfloor r^{\frac{k-1}{2}} \rfloor$ with $r = \sqrt{2}+1$. Note that these are integer boundary values, which is the reason for considering the modified strategy $\hat{\mathbf{w}}$ instead of the original $\mathbf{w}$. From this continuous-valued sequence, we construct the integer-valued *discretized* version $\mathbf{u} = [u_1, u_2, ...]$ by concentrating the probability density of v_k onto integer points, i.e., setting $u_k = \lfloor v_k \rfloor$ for all k. This discretization assigns the following probability mass function to each u_k:

$$Pr(u_k = l) = \int_l^{l+1} f_{v_k}(x)\, dx \quad \text{if } b_k \leq l \leq b_{k+1} - 1 \quad (32)$$

Using this with our strategy $\mathbf{v}$ in (31) gives the following with $b_k = \lfloor (\sqrt{2}+1)^{\frac{k-1}{2}} \rfloor$:

$$Pr(u_k = l) = \frac{2l+1}{{b_{k+1}}^2 - {b_k}^2} \quad \text{if } b_k \leq l \leq b_{k+1} - 1\ . \quad (33)$$

The cost ratio for $\mathbf{u}$ under the quadratic cost function is depicted in Figure 4 (TOP). Note that this plot is numerically very similar to that of optimal uniformly randomized strategy under linear cost. In both cases, the randomized sequences obtain an asymptotic maximum worst-case cost of approximately 2.9142. On the other hand, if the uniform randomization of Definition 4 is applied directly to this boundary sequence under the quadratic cost function, then we obtain the dotted curve in Figure 4 (TOP) which exhibits oscillations, and obtains a maximum cost ratio of roughly 3.06.

Similar methods can be used to obtain strategies for other cost functions. In particular, if $C(\cdot) \in \mathbb{C}$, one can create a continuous uniformly randomized strategy $\hat{\mathbf{w}}$ with the k-th boundary value equal to $C\left(\left\lfloor C^{-1}(r^{k-1}C(1))\right\rfloor\right)/C(1)$. The performance of this strategy under linear cost function will be similar to the optimal uniformly randomized strategy. Then the mapping of (26) can be used to create a strategy $\mathbf{v}$ under cost $C(x)$. Finally, apply the discretization described in (32) to $\mathbf{v}$ to obtain the discrete strategy $\mathbf{u}$, where the k-th TTL random variable will have the following distribution for $b_k \leq l \leq b_{k+1} - 1$:

$$Pr(u_k = l) = \frac{C(l+1) - C(l)}{C(b_{k+1}) - C(b_k)} \quad (34)$$

where $b_k = \left\lfloor C^{-1}(r^{k-1}C(1))\right\rfloor$. Note that while the intermediate step (mapping from $\hat{\mathbf{w}}$ to $\mathbf{v}$) requires $C(\cdot) \in \mathbb{C}$, the final distribution in (34) does not. Therefore this method can be applied when the search cost is only defined for integer values (when $C^{-1}(r^{k-1}C(1))$ is also not defined, b_k can take approximate values). As a result, this method may be more practical than the optimal strategy presented in Section IV. The extent of the similarity between this derived strategy under cost $C(x)$ and the optimal uniformly randomized strategy under linear cost will depend on $C(x)$, due to the fact that we adjusted our boundary values earlier when creating $\hat{\mathbf{w}}$.

B. Discussion

We can calculate the best and worst realization cost ratio, as well as the cost ratio variance, of uniformly randomized strategies in a similar way to that presented in Section IV. Figure 4 (BOTTOM) depicts the performance of the uniformly randomized California Split algorithm under the linear cost function with respect to these metrics. It can be seen from the figure that the worst-case worst-realization cost ratio is 7, much higher than the lower bound of 4. This is because the k-th TTL value is uniformly distributed among all integers between 2^{k-1} and $2^k - 1$, independent of the selection of the previous TTL values, so some inefficient realizations are possible. For example, it is possible for the k-th TTL value to be $2^k - 1$ and the $(k+1)$-th to be 2^k. On the other hand, if successive TTL values are non-independent, then such inefficient realizations may be removed. Figure 5 (TOP) depicts one example of how the probability distribution of the TTL random variables can be jointly defined to decrease the worst-case worst-realization cost ratio while not increasing the worst-case expected cost ratio. Under the randomization proposed by this figure, if the k-th TTL value takes realization $2^{k-1}+\delta$ for some integer $0 \leq \delta \leq 2^{k-1} - 1$, then the $(k+1)$th TTL value will be either $2^k + 2\delta$ with probability $p_{k,\delta+1}$, or it will be $2^k + 2\delta + 1$ with probability $1 - p_{k,\delta+1}$.

Figure 5 (BOTTOM) depicts the cost ratio for this non-independent randomization by setting $p_{i,j} = \frac{1}{2}$ for all i and j. Note that this randomization does not decrease the worst-case cost ratio; however, it does reduce the cost ratio at any non-boundary point (i.e. when $x \neq 2^k$ for all integers k). We see that the worst-case worst-realization cost ratio of this strategy is 4, compared to 7 for the uniformly randomized version. In addition, by comparing Figures 4 (TOP) and 5 (BOTTOM), it can be seen that the cost ratio for the tree construction has less deviation from its mean value.

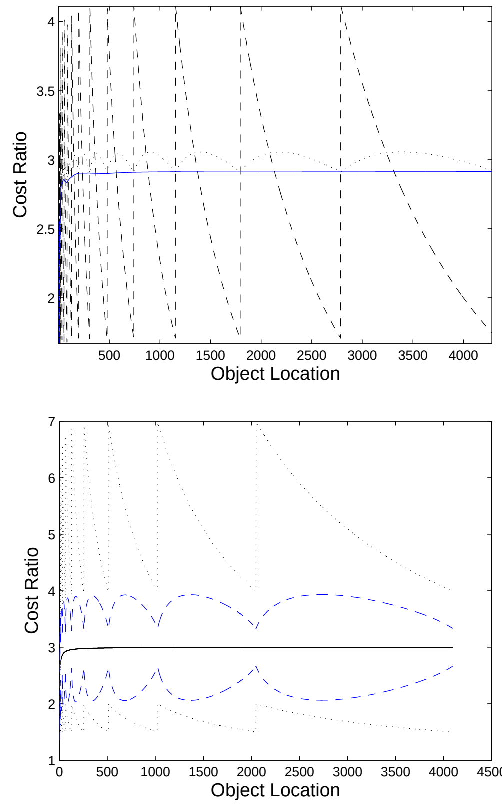


Fig. 4. (TOP): Under a quadratic cost function, the cost ratio as a function of object location for a nonrandom TTL sequence (dashed) with $b_k = \lfloor r^{\frac{k-1}{2}} \rfloor$, $r = \sqrt{2}+1$, its uniformly randomized version (dotted), and its randomized version (solid) corresponding to (33). (BOTTOM): Performance of uniformly randomized California Split rule under a linear cost function. Worst and best realization cost ratio (dotted), average cost ratio (solid), average cost ratio $+/-$ one standard deviation (dashed).

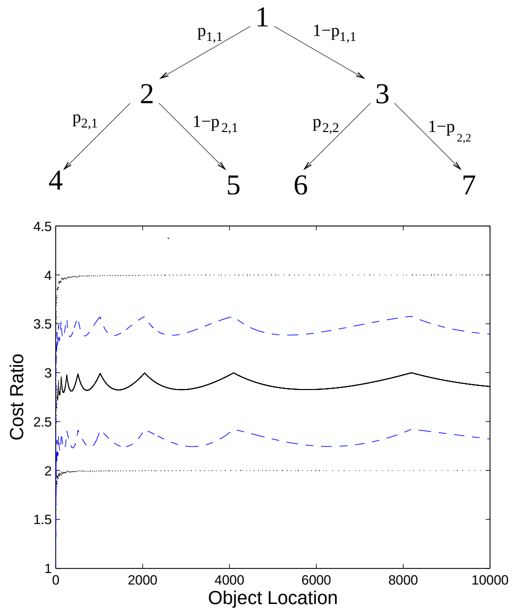


Fig. 5. (TOP): Example of how a binary tree can be used to construct a TTL sequence. In particular, first TTL value is 1. With probability $p_{1,1}$, the second TTL value will either be 2 (with probability $p_{1,1}$) or 3 (with probability $1 - p_{1,1}$), and so on. (BOTTOM): Performance of randomization illustrated in (TOP) figure for $p_{i,j} = \frac{1}{2}$ for all i and j, under a linear cost function. Best and worst realization cost ratio (dotted), average cost ratio (solid) line, and average cost ratio $+/-$ one standard deviation (dashed).

Note that the California Split algorithm was chosen for the tree algorithm only for demonstrative purposes. In fact, for any uniformly randomized strategy, it is possible to use a modified version of the tree construction given by Figure 5 (TOP) to obtain the same value of worst-case cost ratio but with lower worst-case worst-realization cost ratio. The tree construction is modified by adjusting the number of nodes in each level of the tree, and modifying the transition probabilities from nodes in successive levels.

VIII. Conclusion

In this paper we studied the class of TTL-based controlled flooding search methods used to locate an object/node in a large network. When the object location distribution is not known and adopting a worst-case performance measure, we showed that *randomized* search strategies outperform fixed strategies. We derived an asymptotically optimal strategy whose search cost is always within a factor of e of the cost of an omniscient observer. We also studied the optimal strategy under alternative performance measures. We further provided a mapping between TTL sequences under different cost functions, and investigated the class of uniformly randomized strategies. These results are directly applicable to the design of practical algorithms.

Appendix

A. Proof of Lemma 5

First note that because $C(\cdot) \in \mathbb{C}$, $v_{j+1} > v_j$ if and only if $C(v_{j+1}) > C(v_j)$. From (13), the ratio between cost of successive TTL values can be expressed in terms of $C(v_1)$ as follows for any integer $j \geq 1$:

$$\frac{C(v_{j+1})}{C(v_j)} = \left(\alpha^{-\sum_{k=0}^{j-1}\alpha^k}\right)\left(\frac{C(v_1)}{C(1)}\right)^{\alpha^j} \tag{35}$$

Consider any positive finite j. If (16) holds then we have by using (35):

$$\begin{aligned}\frac{C(v_{j+1})}{C(v_j)} &\geq \left(\alpha^{-\sum_{k=0}^{j-1}\alpha^k}\right)\left(\alpha^{\left(\sum_{k=1}^{\infty}\alpha^{-k}\right)}\right)^{\alpha^j} \\ &> \left(\alpha^{-\sum_{k=0}^{j-1}\alpha^k}\right)\left(\alpha^{\left(\sum_{k=1}^{j}\alpha^{-k}\right)}\right)^{\alpha^j} = 1\ ,\end{aligned}$$

which holds for all integers j. Hence, (16) is a sufficient condition for $\mathbf{v}$ to be increasing.

Now suppose $\mathbf{v}$ is increasing. Then for any positive integer j we have by rearranging (35) and using $C(v_{j+1}) > C(v_j)$:

$$\begin{aligned}\frac{C(v_1)}{C(1)} &= \left[\frac{C(v_{j+1})}{C(v_j)}\left(\alpha^{\sum_{k=0}^{j-1}\alpha^k}\right)\right]^{\alpha^{-j}} \\ &> \alpha^{\alpha^{-j}\sum_{k=0}^{j-1}\alpha^k} = \alpha^{\sum_{k=1}^{j}\alpha^{-k}}\end{aligned}$$

Taking the limit of this inequality as j approaches ∞ gives: $\frac{C(v_1)}{C(1)} \geq \alpha^{\left(\sum_{k=1}^{\infty}\alpha^{-k}\right)} = \alpha^{\frac{1}{\alpha-1}}$, thereby proving that (16) is also a necessary condition for an increasing sequence. ∎

B. Proof of Lemma 7

First note that from (J.2), we have that $C(v_k) = r^{k-1}C(v_1)$ for all $k \geq 1$. Let $S_k = C(v_1)+C(v_2)+....+C(v_k)$ for $k \geq 1$. The expected value of S_k can be calculated as follows:

$$\begin{aligned}E\left[S_k\right] &= E\left[\sum_{j=1}^{k} C(v_j)\right] = E\left[\sum_{j=1}^{k} r^{j-1}C(v_1)\right] \\ &= \sum_{j=0}^{k-1} r^j E\left[C(v_1)\right] = E\left[C(v_1)\right]\frac{r^k-1}{r-1}\end{aligned} \tag{36}$$

In addition, the conditional expectation of $C(v_1)$ can be calculated as follows, for $1 \leq l < C^{-1}(r \cdot C(1))$:

$$\begin{aligned}E\left[C(v_1)|v_1 \leq l\right] &= \int_0^{\infty} Pr\left(C(v_1) > y | v_1 \leq l\right)\ dy \\ &= C(1) + \int_{C(1)}^{\infty} \frac{Pr\left(C^{-1}(y) < v_1 \leq l\right)}{Pr(v_1 \leq l)}\ dy \\ &= C(1) + \frac{1}{F_{v_1}(l)}\left[\int_{C(1)}^{C(l)}\left[\bar{F}_{v_1}(C^{-1}(y)) - \bar{F}_{v_1}(l)\right]\ dy\right] \\ &= \frac{1}{F_{v_1}(l)}\left[h\left(\frac{C(l)}{C(1)}\right) - C(l)\cdot\bar{F}_{v_1}(l)\right]\ .\end{aligned} \tag{37}$$

We will use the following notation. $J_x^{\mathbf{v}}|_{v_n > x}$ denotes the conditional expected search cost of using strategy $\mathbf{v}$ when the object location is x, given that $v_n > x$. Similarly, $J_x^{\mathbf{v}}|_{v_n \leq x}$ is the conditional expected search cost given that $v_n \leq x$.

Now consider any real number $x \geq 1$; there must exist a positive integer n such that $r^{n-1}C(1) \leq C(x) < r^n C(1)$, or equivalently $C^{-1}\left(r^{n-1}C(1)\right) \leq x < C^{-1}\left(r^{n}C(1)\right)$. Then by using (36), the expected search cost $J_x^{\mathbf{v}}$ can be calculated:

$$\begin{aligned}J_x^{\mathbf{v}} &= J_x^{\mathbf{v}}|_{v_n > x} Pr\left(v_n > x\right) + J_x^{\mathbf{v}}|_{v_n \leq x} Pr\left(v_n \leq x\right)\\ &= E\left[S_n | v_n > x\right] Pr\left(v_n > x\right) + E\left[S_{n+1} | v_n \leq x\right] Pr\left(v_n \leq x\right)\\ &= E[S_n] + E[C(v_{n+1}) | v_n \leq x] Pr(v_n \leq x)\\ &= \frac{r^n - 1}{r-1} E\left[C(v_1)\right]\\ &\quad + r^n E\left[C(v_1) \middle| v_1 \leq C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right] F_{v_1}\left(C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right)\end{aligned}$$

Using (37), we obtain the following:

$$\begin{aligned}J_x^{\mathbf{v}} &= \frac{r^n - 1}{r-1} E\left[C(v_1)\right]\\ &\quad + r^n \left[h\left(\frac{C(x)}{r^{n-1}C(1)}\right) - \frac{C(x)}{r^{n-1}} \bar{F}_{v_1}\left(C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right)\right]\\ &= \frac{r^n}{r-1}\left[E\left[C(v_1)\right] + (r-1) h\left(\frac{C(x)}{r^{n-1}C(1)}\right)\right]\\ &\quad - r C(x) \bar{F}_{v_1}\left(C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right) - \frac{E\left[C(v_1)\right]}{r-1}\ . \qquad (38)\end{aligned}$$

Letting $z = \frac{C(x)}{r^{n-1}C(1)}$, we obtain the following expression for the cost ratio by plugging into (38):

$$\frac{J_x^{\mathbf{v}}}{C(x)} = \frac{r}{r-1}\frac{h(r) + (r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} \qquad (39)$$

$$- \frac{h(r)}{(r-1) z r^{n-1} C(1)}\ , \qquad (40)$$

where we have used the fact that $h(r) = E\left[C(v_1)\right]$ (by the relationship between expectation and tail distribution), and $h'(z) = \bar{F}_{v_1}\left(C^{-1}(z \cdot C(1))\right) \cdot C(1)$ from basic calculus.

For simplicity of notation, define $\Phi_n(z)$ as equal to the right-hand side of (39) and (40), so that $\Phi_n(z)$ is simply the cost ratio at object location $x = C^{-1}\left(z r^n C(1)\right)$. The following is true for any x and $y = C^{-1}(rC(x))$: $\frac{J_x^{\mathbf{v}}}{C(x)} < \frac{J_y^{\mathbf{v}}}{C(y)}$. This statement holds because the two terms on the right-hand side of (39) are the same for x and y, and the term in (40) increases with increasing x. In addition, when x ranges from $C^{-1}(r^{n-1}C(1))$ to $C^{-1}(r^n C(1))$, then z takes values between 1 and r. Hence, we have $\Phi_n(z) < \Phi_{n+1}(z)$ for all n and z. Finally, note that the limit as $n \to \infty$ of $\Phi_n(z)$ is simply $\Phi(z)$, where $\Phi(z)$ is the function defined earlier in (19). Hence, the following is true, where $x_n = C^{-1}(r^n C(1))$:

$$\begin{aligned}\sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} &= \sup_{n \in \mathbb{Z}^+}\left\{\sup_{x_{n-1} \leq x < x_n} \frac{J_x^{\mathbf{v}}}{C(x)}\right\}\\ &= \sup_{n \in \mathbb{Z}^+}\left\{\sup_{1 \leq z < r} \Phi_n(z)\right\} = \sup_{1 \leq z < r}\left\{\sup_{n \in \mathbb{Z}^+} \Phi_n(z)\right\}\\ &= \sup_{1 \leq z < r}\left\{\lim_{n \to \infty} \Phi_n(z)\right\} = \sup_{1 \leq z < r}\left\{\Phi(z)\right\}\\ &= \sup_{1 \leq z < r}\left\{\frac{r}{r-1}\frac{h(r) + (r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)}\right\}\ ,\end{aligned}$$

which completes the proof of the lemma. ∎

C. *Proof of Proposition 1*

First we examine the worst-case realization cost ratio for general strategies of the type $\mathbf{v}[r, F_{v_1}(x)]$. Fix the object location x. There must exist exactly one k such that $C^{-1}\left(r^{k-1}C(1)\right) \leq x < C^{-1}\left(r^{k}C(1)\right)$. Note that the particular realization of the sequence $\mathbf{v}$ is uniquely defined by the realization of the first TTL random variable v_1. Let $\tilde{\mathbf{v}} = [\tilde{v}_1, \tilde{v}_2, ...]$ denote a realization of $\mathbf{v}$. The worst-realization cost ratio is when $\tilde{v}_1$ approaches $C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)$ from below. This is true because at these values, the k-th TTL value is slightly less than x and hence the $(k+1)$-th TTL value will be needed to complete the search. The worst-realization cost ratio is thus upper bounded by:

$$\begin{aligned}\Gamma_x^{\mathbf{v}} &\leq \lim_{\tilde{v}_1 \to \left(C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right)^-}\left\{\frac{1}{C(x)}\sum_{j=1}^{k+1} r^{j-1} C\left(\tilde{v}_1\right)\right\}\\ &= \frac{r^{k+1} - 1}{(r-1) r^{k-1}} = \frac{r^2 - r^{-k+1}}{r-1}\ . \qquad (41)\end{aligned}$$

This bound increases as x increases, and it easily follows that:

$$\Gamma^{\mathbf{v}} \leq \lim_{x \to \infty} \Gamma_x^{\mathbf{v}} \leq \lim_{k \to \infty} \frac{r^2 - r^{-k+1}}{r-1} = \frac{r^2}{r-1}\ , \qquad (42)$$

The inequality above becomes equality when the probability density function for v_1 is strictly positive in the interval $\left[C^{-1}\left(\frac{C(x)}{r^{k-1}}\right) - \epsilon, C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right)$, for some $\epsilon > 0$. This is true because if the density function for v_1 is positive in this interval, then there is a nonzero probability that v_1 is arbitrarily close to $C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)$. Then all of the inequalities in (41) and (42) become equalities. This condition on the density function is satisfied when $F_{v_1}(x) = \frac{1}{\ln r}\ln\frac{C(x)}{C(1)}$, and hence strategies with this family of cdf have worst-case worst-realization cost ratio value of $r^2/(r-1)$.

Similarly, the best-realization cost ratio of these types of strategies can be calculated for object location x, where $C^{-1}\left(r^{k-1}C(1)\right) \leq x < C^{-1}\left(r^{k}C(1)\right)$. It can be easily shown that the best-case realization occurs when $\tilde{v}_1$ is such that $C(\tilde{v}_1) = C(x)/r^{k-1}$. In this case, the cost ratio can be calculated as: $\gamma_x^{\mathbf{v}} = \frac{\sum_{j=1}^{k} r^{j-1}\frac{C(x)}{r^{k-1}}}{C(x)} = \frac{r - r^{-k+1}}{r-1}$, which approaches $\frac{r}{r-1}$ as $x \to \infty$.

We now examine the cost ratio variance of strategy $\mathbf{v}[r,\ F_{v_1}(x)]$ where $F_{v_1}(x) = (\ln C(x)/C(1))/\ln r$. Let $f_{v_1}\left(x \,|v_1 < y\right)$ denote the pdf of v_1 given that v_1 is less than y. Then for $1 \leq y < C^{-1}\left(rC(1)\right)$,

$$\begin{aligned}f_{v_1}\left(x \,|v_1 < y\right) &= \frac{dF_{v_1}\left(x \,|v_1 < y\right)}{dx}\\ &= \begin{cases}\frac{dC(x)/dx}{C(x)\ \ln\left[C(y)/C(1)\right]} & \text{if} \quad 1 \leq x < y\\ 0 & \text{otherwise}\end{cases}\end{aligned}$$

Then we have:

$$\begin{aligned}E\left[C\left(v_1\right)^2 \middle| v_1 < y\right] &= \int_1^y \frac{C(x)^2}{C(x)\ln\left[C(y)/C(1)\right]}\frac{dC(x)}{dx}\ dx\\ &= \frac{C(y)^2 - C(1)^2}{2\ln\left[C(y)/C(1)\right]} \qquad (43)\end{aligned}$$

On the other hand, since $E\left[C\left(v_1\right)^2\right] = \frac{C(1)^2}{\ln r}\left[\frac{r^2-1}{2}\right]$, then we obtain the following:

$$\begin{aligned} &E\left[C\left(v_1\right)^2 \middle| v_1 \geq y\right] \\ &= \frac{E\left[C(v_1)^2\right] - E\left[C\left(v_1\right)^2 \middle| v_1 < y\right] Pr\left(v_1 < y\right)}{Pr(v_1 \geq y)} \\ &= \frac{C(1)^2 r^2 - C(y)^2}{2\left(\ln r - \ln\left[C(y)/C(1)\right]\right)} \end{aligned} \quad (44)$$

Now fix any x; there must exist a positive integer k such that $C^{-1}\left(r^{k-1}C(1)\right) < x \leq C^{-1}\left(r^k C(1)\right)$. As defined earlier, we let $j_x^{\mathbf{v}}$ be a random variable denoting the cost of using strategy $\mathbf{v}$ when object location is x. Note that $J_x^{\mathbf{v}} = E\left[j_x^{\mathbf{v}}\right]$. For notation, let $F_{v_k}(x) = Pr(v_k < x)$ and $\bar{F}_{v_k}(x) = Pr(v_k \geq x)$ for any integer k and location x. Also, let $x_k = C^{-1}(C(x)/r^{k-1})$. The second moment of the search cost can be calculated as follows by using (43) and (44):

$$\begin{aligned} E\left[(j_x^{\mathbf{v}})^2\right] &= E\left[\left(\sum_{l=1}^{k+1} r^{l-1}C(v_1)\right)^2 \middle| v_k < x\right] F_{v_k}(x) \\ &\quad + E\left[\left(\sum_{l=1}^{k} r^{l-1}C(v_1)\right)^2 \middle| v_k \geq x\right] \bar{F}_{v_k}(x) \\ &= \left(\frac{r^{k+1}-1}{r-1}\right)^2 E\left[C(v_1)^2 \left| v_1 < x_{k-1}\right.\right] F_{v_1}\left(x_{k-1}\right) \\ &\quad + \left(\frac{r^{k}-1}{r-1}\right)^2 E\left[C(v_1)^2 \left| v_1 \geq x_{k-1}\right.\right] \bar{F}_{v_1}\left(x_{k-1}\right) \\ &= \frac{A \cdot C(x)^2}{2(\ln r)(r-1)^2}\ , \end{aligned} \quad (45)$$

where A is defined as follows:

$$\begin{aligned} A =& (r^{k+1}-1)^2\left[r^{-2k+2} - \frac{C(1)^2}{C(x)^2}\right] \\ &+ (r^k-1)^2\left[\frac{C(1)^2}{C(x)^2}r^2 - r^{-2k+2}\right]\ . \end{aligned}$$

In addition, it can be easily shown that $J_x^{\mathbf{v}} = \frac{rC(x)-C(1)}{\ln r}$. The cost ratio variance at location x is simply the difference between (45) and $(J_x^{\mathbf{v}})^2$, divided by $C(x)^2$. Hence we have after combining terms:

$$\frac{\Lambda_x^{\mathbf{v}}}{C(x)^2} = \frac{A}{2(\ln r)(r-1)^2} - \left(\frac{r - \frac{C(1)}{C(x)}}{\ln r}\right)^2 .$$

Note that as x approaches ∞ (so that $C(x)$ also approaches infinity), the cost ratio variance becomes:

$$\begin{aligned} &\lim_{x\to\infty} \frac{\Lambda_x^{\mathbf{v}}}{C(x)^2} \\ &= \lim_{k\to\infty} \frac{(r^{k+1}-1)^2\left[r^{-2k+2}\right] - (r^k-1)^2 r^{-2k+2}}{2(\ln r)(r-1)^2} - \left(\frac{r}{\ln r}\right)^2 \\ &= \frac{r^4 - r^2}{2(\ln r)(r-1)^2} - \frac{r^2}{(\ln r)^2} \end{aligned}$$

■

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their thoughtful comments.

REFERENCES

[1] D. Braginsky and D.Estrin, "Rumor routing algorithm for sensor networks," *Proc. International Conference on Distributed Computing Systems (ICDCS-22)*, 2002.

[2] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, pp. 153–181, 1999.

[3] J. Xie, R. Talpade, T. McAuley, and M. Liu, "AMRoute: Ad Hoc multicast routing protocol," *ACM Mobile Networks and Applications (MONET) Special Issue on Mobility of Systems, Users, Data and Computing*, vol. 7, no. 6, pp. 429–439, December 2002.

[4] C. Carter, S. Yi, P. Ratanchandani, and R. Kravets., "Manycast: exploring the space between anycast and multicast in ad hoc networks," *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networks (MobiCOM'03)*, September 2003, San Diego, California.

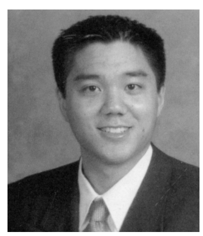
[5] N. Chang and M. Liu, "Revisiting the TTL-Based controlled flooding search: optimality and randomization," *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networks (MobiCOM'04)*, pp. 85–99, September 2004, Philadelphia, PA.

[6] Y. Baryshinikov, E. Coffman, P. Jelenkovic, P. Momcilovic, and D. Rubenstein, "Flood search under the california split rule," *Operations Research Letters*, vol. 32, no. 3, pp. 199–206, May 2004.

[7] S. Ni, Y. Tseng, Y. Chen, and J. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networks (MobiCOM'99)*, August 1999, Seattle, Washington.

[8] A. Borodin and R. El-Yaniv, *Online Comptutation and Competitive Analysis*, Cambridge University Press, 1998, Cambridge, UK.

[9] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, 1996, New York, USA.



Nicholas B. Chang received the B.S.E. degree in electrical engineering (*magna cum laude*) from Princeton University, Princeton, NJ in 2002, and the M.S.E degree in electrical engineering: systems and M.S. degree in mathematics from the University of Michigan, Ann Arbor, in 2004 and 2005, respectively.

He is currently a Ph.D. candidate in the Electrical Engineering: Systems Department at the University of Michigan, Ann Arbor. His research interests include communication networks, wireless communication, stochastic control, stochastic resource allocation, and algorithms. During the summers of 2001 and 2002, he worked for MIT Lincoln Laboratory in Lexington, MA.

Mr. Chang is a Member of Tau Beta Pi and Sigma Xi, the Scientific Research Society. He is a recipient of the MIT Lincoln Laboratory Graduate Fellowship for the 2005-06 academic year.

Mingyan Liu (M'00 / ACM'01) received her B.Sc. degree in electrical engineering in 1995 from the Nanjing University of Aero. and Astro., Nanjing, China, M.Sc. degree in systems engineering and Ph.D. Degree in electrical engineering from the University of Maryland, College Park, in 1997 and 2000, respectively. She joined the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, in September 2000, where she is currently an Assistant Professor. Her research interests are in performance modeling, analysis, energy-efficiency and resource allocation issues in wireless mobile ad hoc networks, wireless sensor networks, and terrestrial satellite hybrid networks. She is the recipient of the 2002 NSF CAREER Award, and the University of Michigan Elizabeth C. Crosby Research Award in 2003.