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 also show that in the finite case, for a given deterministic TTL sequence there exists a randomized version that attains a lower worst-case search cost. We further introduce a class of simple (sub-optimal but potentially more useful in practice) randomized strategies and derive the optimal strategy within this class.

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 network. The ability to conduct cost effective and fast searches has become an increasingly critical component required by many emerging networks and applications. A prime example is data query in a large wireless sensor network, where different data is distributed among a large number of sensor nodes based on different sensor readings. A query may be initiated by any node in search of certain data of interest (e.g., the position coordinates where temperature has exceeded a certain level) [1]. As it is not known *a priori* where the data might be located, or which node has the data, the query has to be somehow advertised to nodes in the network. As the query propagates, a node that has data matching the interest will respond to the querying node with the desired data [2]. There may be more than one node in the network (or sometimes none) that has the queried data. Depending on the underlying application, we may need to locate one, some, or all of these nodes. Search has also been extensively used in mobile ad hoc networks. This includes searching for a destination node by a source node in the route establishment procedure of an ad hoc routing protocol (e.g., [3]), searching for a multicast group by a node looking to join the group (e.g., [4]), and locating one or multiple servers by a node requesting distributed services (e.g., [5]). Search is also widely used in peer-to-peer (P2P) networks, marked by the need to locate desired objects/files that are shared among nodes in the network.

There are a variety of mechanisms one may use to search/locate a node or object in a large network. The first is to maintain a centralized directory service, where nodes issue queries to the central directory to obtain the location of the search target. The central directory needs to be constantly updated as the network topology and data content change. Such systems tend to have very short response time, if the directory information is kept afresh. On the other hand, centralized systems often scale poorly as the network increases in size, and as location information changes more frequently (either due to topology change as a result of mobility or due to the information content change in the network). The latter necessitates a large amount of information update which can cause significant energy consumption overhead, especially when the queries occur less frequently compared to changes in the network.

A second class of methods, which is decentralized, is the random walk based search, where the querier sends out a query packet which is forwarded in a random walk fashion, until it hits the search target. These can be pure random walks or controlled walks such that the propagation of the packet is maintained in an approximately consistent direction. In particular, [2] proposed random walks initiated by both the querier and the node that has data of potential interest (called advertisement). There have been many results on estimating the search cost and response time using such approaches, see for example [6].

In this paper we focus on a widely used search mechanism known as the TTL-based controlled flooding of query packets. This is also a decentralized approach. 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, the corresponding node 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 (e.g., in terms of energy consumption in a wireless network associated with the amount of packet transmissions/receptions). These strategies may be applied to wired and wireless networks alike, although decentralized and unstructured searches are more relevant in a wireless scenario, particularly in a wireless sensor network. We will not explicitly consider the response time of a search strategy. One reason is that within the class of controlled flooding search, the fastest search is to flood the entire network. In addition, if the search cost is a function of the number of packet transmissions and receptions, then the goal of minimizing cost is loosely aligned with the objective of locating the object quickly. We will limit our analysis to the case of searching for a single target, which is assumed to exist in the network. 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 searching. We will use the term *location* of an object to indicate both the actual position of the search target and the minimum TTL value needed to locate this object.

When the probability distribution of the location of the object is known *a priori*, search strategies that minimize the expected search cost can be obtained via a dynamic programming formulation [7]. The necessary and sufficient conditions were also derived in [7] for two very commonly used search strategies to be optimal. When the distribution of the object location is not known *a priori*, one may evaluate the effectiveness of a strategy by its worst case performance. In [8] 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 and 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 this paper we show that *randomized strategies* perform better for the same criterion. 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.

The main results of this paper are summarized as follows.

1) We show that given a deterministic TTL sequence, there exists a randomized version that has a lower worst-case expected search cost. The construction of the randomization is presented.
2) We derive an asymptotically (as the network size increases) optimal strategy and show that its worst-case cost ratio is e.
3) We establish an equivalence between TTL sequences under different cost functions. This allows us to use results for sequences derived under specific cost functions in order to derive strategies for other cost functions.
4) We introduce a class of *uniformly randomized strategies* and showed that within this class the best strategy achieves a competitive ratio of 2.9142. Though sub-optimal, these strategies are simple to implement and of practical value.

The rest of the paper is organized as follows. In Section II we present the network model, assumptions, and the search cost function as an abstraction of lower layer networking mechanisms. In Section III we introduce the performance objectives under consideration as well as some preliminary results. In Section IV we show how a randomized strategy may be constructed given a finite non-random strategy to result in better worst-case performance. In Section V we derive the optimal strategy among all random and non-random strategies with respect to the performance criterion given in the next section. This is done for continuous and discrete strategies, respectively. We also examine a few alternative performance measures in Section VI. In Sections VII and VIII 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. In particular we establish an equivalence relationship between a linear cost function and more general cost functions in Section VII. Using this result in Section VIII we introduce a class of strategies called *uniformly randomized strategies* and derive the optimal strategy within this class. Finally Section IX concludes the paper.

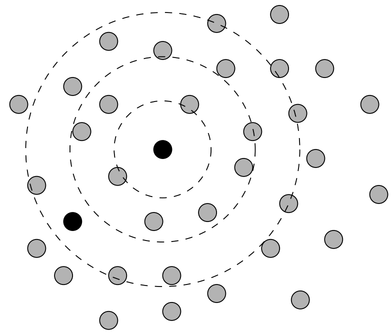

Fig. 1. Example network in which TTL-based controlled flooding search is employed. In the figure, the center black node originates the search by passing a TTL value and search query to its one-hop neighbors. The process is repeated until either TTL value reaches 0 or object is found. Note that using a TTL value of k will approximately reach all neighbors within a circle of radius of $k \cdot r$, where r is the single-transmission radius. In this diagram, object is located in the left black node and hence using a TTL value of 3 or larger will complete the search.

II. Network Model

A. 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. In particular, in a wireless scenario each transmission covers a specific region given the limitation on the transmission power, channel fading, etc. All nodes within that region will be considered one-hop neighbors of the transmitting node.

The node originating the search begins by determining an initial positive integer TTL value, and passes this number along with its search query to its neighboring nodes. If the underlying network is wired, this query will be transmitted once along each outgoing link of the originating node. For a wireless network, the originating node can reach all its neighbors in a single broadcast transmission. If the object is found at a neighboring node, then the corresponding node will reply to the originating node. If a neighboring node does not have the desired object, it will decrement the TTL value by one and pass the query to its neighbors in the same fashion. In this way the query packets are duplicated and propagated in the network. The above process repeats until either the object has been located or the TTL value reaches 0, at which point the query packet is dropped. This process is depicted in Figure 1 for a two-dimensional network, in which the middle black source node uses a TTL value of 3 in order to locate the desired object. The originating node starts a timer when the first query packet is sent. If it does not get a response back before the timer expires, it will begin a new *round* of search by selecting a TTL of an increasing value, and the above procedure is repeated. The TTL value is increased in subsequent rounds until the object is located.

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 [9]. For example, a node should suppress multiple copies of the same query it receives. In our analysis we will ignore these technical details, and simply 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. For the rest of our discussion we will no longer regard network as wired or wireless, but only discuss in terms of the search cost $C(k)$, since in essence it abstracts the relevant features of lower layers.

We summarize the assumptions underlying our network model as follows.

1) We assume that a single target object exists in the network. Therefore flooding the entire network will for sure locate the object.
2) We assume that a TTL value of k will reach all nodes within k hops of the originating node when the timer expires.
3) A search with TTL value of k incurs a cost $C(k)$. The functional form of this cost depends on the properties of the network as well as the underlying broadcast techniques mentioned earlier.

Assumption (2) is a simplification. In particular, in a wireless network this assumption implies that the redundancy inherent in the query broadcast process ensures that a node receives correctly at least one copy of the same query in spite of possible packet collisions and channel interference. It also implies that there is no excessive delay in the network, thus a timeout event is equivalent to not finding the object in the k-hop neighborhood. This simplication nevertheless allows us to reveal fundamental features of the problem and obtain insights.

B. Search Strategies

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. We define the following classes of search strategies.

Definition 1: A deterministic integer-valued search strategy $\mathbf{u}$ is a sequence $\mathbf{u} = [u_1, u_2, \cdots, u_N]$ for some N, where u_i is a positive integer and is the TTL value used during the i-th round of search. In addition, $u_i < u_{i+1}$ for all $1 \leq i \leq N-1$.
The requirement for the sequence to be increasing is a natural one. Note that in a specific search experiment we may not need to use the entire sequence. We will also refer to deterministic strategies as *fixed* or *non-random*.

Definition 2: A randomized discrete search strategy $\mathbf{u}$ is a TTL sequence that consists of random variables that take on integer values, i.e., $\mathbf{u} = [u_1, u_2, \cdots, u_N]$, where u_i is a discrete random variable that takes integer values, $1 \leq i \leq L$. The distribution of u_i can be independently or jointly defined. This is a straightforward generalization of the previous definition.

For analyzing TTL-based controlled flooding search, it is natural to only consider integer-valued policies. However, as will become evident later, considering *real-valued* sequences can reveal fundamental properties that will be helpful in deriving optimal integer-valued strategies. We define continuous/real-valued fixed and randomized stratetied in a similar way, respectively.

Definition 3: A deterministic continuous/real-valued search strategy $\mathbf{v}$ is a sequence $\mathbf{v} = [v_1, v_2, \cdots, v_N]$ for some N, where $\leq v_i$ takes any real value on $[1, \infty)$ and is the TTL value used during the i-th round. In addition, $v_i < v_{i+1}$ for all $1 \leq i \leq N-1$.

Definition 4: A randomized continuous/real-valued search strategy $\mathbf{v}$ is a TTL sequence that consists of continuous random variables, i.e., $\mathbf{v} = [v_1, v_2, \cdots, v_N]$, where v_i is a continuous random variable that takes any real value on $[1, \infty)$, $1 \leq i \leq L$. The distribution of v_i can be independently or jointly defined.

Lower bounding the TTL values by 1 in the above definitions allows us to derive a positive integer-valued sequence from any real-valued sequence by simply taking the floor of the latter.

Definition 5: A search strategy is admissible if by using the strategy an object of finite location is located with probability 1. Specifically, given a network of finite dimension L, a deterministic strategy $\mathbf{u}$ of length N is admissible if $u_N = L$, and a randomized (discrete or continuous) strategy $\mathbf{u}$ of length N is admissible if $Pr(u_n = L) = 1$ for some $1 \leq n \leq N$. In the asymptotic case where $L \Rightarrow \infty$ and infinite TTL sequences are employed, a strategy $\mathbf{u} = [u_1, u_2, \cdots]$ is admissible if for $\forall\, x \geq 1,\ \exists\, n \in \mathbb{Z}^+$ s.t. $Pr\,(u_n \geq x) = 1$.

For the rest of this paper we will let V denote the set of all real-valued admissible strategies (random or fixed). Any strategy $\mathbf{v} \in V$ will be referred to as a *continuous* or *real-valued* strategy. 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). Any strategy $\mathbf{u} \in U$ will be referred to as a *discrete* or *integer-valued* strategy. Finally, U^d denotes the set of all admissible integer-valued deterministic strategies. Note that $U^d \subset U \subset V$ and $U^d \subset V^d \subset V$.

C. Object Location and Search Costs

A search cost $C(v)$ is incurred by a round of search with TTL value v. 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.

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 is defined for positive integers v. We will use the same function $C(v)$ in both situations, as it will be clear from the context whether C needs to be defined for all integers or real x.

We will adopt the natural assumption that $C(v_1) > C(v_2)$ if $v_1 > v_2$, i.e., the cost strictly increases as the search covers a bigger region. We define the following class of cost functions for real-valued sequences.

Definition 6: The function $C : [1, \infty) \rightarrow [C(1), \infty)$ belongs to the class $\mathbb{C}$ if $0 < C(1) < \infty$, $C(v)$ is increasing and differentiable (hence continuous), and $\lim_{v \rightarrow \infty} C(v) = \infty$. Note that for every $y \in [C(1), \infty)$, there exists exactly one $v \in [1, \infty)$ such that $C(v) = y$.

In our discussion and numerical examples we will also use two special cost functions, a linear cost and a quadratic cost, defined as $C(v) = \alpha v$ and $C(v) = \alpha v^2$, respectively, for some constant $\alpha > 0$. The first is a good model in a network where the number of transmissions incurred by the search query is proportional to the TTL value used, e.g., 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 v hops is on the order of v^2. Note that both the linear and quadratic costs are within the class $\mathbb{C}$ when defined over $[1, \infty)$.

We will use X to denote the minimum TTL value required to locate the object. We will refer to X as the object "location". When considering discrete strategies $\mathbf{u} \in U$, we will also assume that X is an integer-valued random variable taking values between 1 and L such that $Pr(X \in \{1, 2, \cdots, L-1, L\}) = 1$. When considering continuous strategies $\mathbf{v} \in V$, we allow the location to take any real value in the interval $[1, L]$. In both cases, we denote the cumulative distribution of X by $F(x)$, where $F(x) = Pr(X \leq x)$. The tail distribution of X is denoted by $\bar{F}(x) = 1 - F(x) = Pr(X > x)$. Note that $F(L) = 1$ and $\bar{F}(L) = 0$ for any X.

III. Problem Formulation and Preliminaries

In this section, we present the performance objective, problem formulation, as well as some preliminaries on key properties of strategies under consideration.

We adopt the following performance measure that reflects the search cost in the worst case scenario (a generalization of the one used in [8]):

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

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 (expected) *cost ratio*. $\{p_X(x)\}$ 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*. $J_X^{\mathbf{u}}$ can be calculated as follows:

$$\begin{aligned} J_X^{\mathbf{u}} &= E_{\mathbf{u}}\left[E_X\left[j_X^{\mathbf{u}}\right]\right] \\ &= E_{\mathbf{u}}\left[\sum_{u_i \in \mathbf{u}} C(u_i) Pr(X > u_{i-1})\right] \ , \end{aligned} \tag{2}$$

where $u_0 = 0$ is assumed for all $\mathbf{u}$ and $j_X^{\mathbf{u}}$ is the random variable denoting the search cost of using strategy $\mathbf{u}$ when object location is X. 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 $\mathbf{u}$. Note that we will only consider the case where the random vector $\mathbf{u}$ and X are mutually independent since the distribution of X is not known *a priori*.

The worst-case cost ratio $\rho^{\mathbf{u}}$ can also be viewed as the *competitive ratio* with respect to an *oblivious adversary* [10] 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 slightly different meanings 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 noted before. In this case, the search cost of using $\mathbf{u}$ is always within a factor $\rho^{\mathbf{u}}$ of the omniscient observer cost for any given location. On the other hand, when $\mathbf{u}$ is random, $\rho^{\mathbf{u}}$ only provides an upper bound on the average search cost but does not necessarily upper bound any particular realization of this cost, as $J_X^{\mathbf{u}}$ is a double expectation with respect to both the strategy and the location. 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 VI, we will present other performance measures in order to account for these differences.

The corresponding objective is to find search strategies that minimize this ratio, denoted by $\mathbf{u}^*$:

$$\rho^* = \inf_{\mathbf{u} \in U} \rho^{\mathbf{u}} = \inf_{\mathbf{u} \in U} \sup_{\{p_X(x)\}} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \ . \tag{3}$$

Note that the quantity $\frac{J_X^{\mathbf{u}}}{E[C(X)]}$ is lower bounded by 1 for all distributions X, hence $\rho^* \geq 1$, since $j_X^{\mathbf{u}} \geq C(X)$ with probability 1 for all X, i.e, any strategy $\mathbf{u}$ cannot outperform an omniscient observer who knows the object location in advance.

For any continuous strategy, $\mathbf{v} \in V$, the worst-case cost ratio is similarly defined as in (1):

$$\rho^{\mathbf{v}} = \sup_{\{f_X(x)\}} \frac{J_X^{\mathbf{v}}}{E[C(X)]} \ , \tag{4}$$

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 similarly defined as in (3) with $\{f_X(x)\}$ replacing $\{p_X(x)\}$.

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 object location density is $f_Y(y') = \delta(y' - y)$, i.e., a single fixed point.

Proof: We begin by noting that for every $y \in [1, \infty)$, there corresponds a singleton probability density $f_Y(y') = \delta(y' - y)$, such that $E[C(Y)] = C(y)$ 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:

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

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

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

Inequalities (6) and (8) collectively imply the equality in equation (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 (3) as

$$\rho^* = \inf_{\mathbf{u} \in U} \rho^{\mathbf{u}} = \inf_{\mathbf{u} \in U} \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} , \tag{10}$$

and similarly for the continuous strategies.

It has been shown in [8] that under a linear cost function $C(u) = \alpha \cdot u$ for some constant α, and as the network size increases asymptotically, the minimum worst-case cost ratio over all deterministic integer-valued sequences is 4, achieved by the California Split search $\bar{\mathbf{u}} = \left\{2^{i-1} : i \in \mathbb{Z}^+\right\} = [1, 2, 4, 8, ...]$.

In the next section we show that it is in fact always possible to find a random TTL sequence for a given finite nonrandom sequence that performs better in the worst case. Therefore, under this criterion the best search strategies are randomized strategies. In Section V we further derive the optimal randomized strategies that are optimal among *all* admissible strategies, which achieve a much smaller worst-case cost ratio (e as opposed to 4) for any cost function $C(\cdot) \in \mathbb{C}$.

IV. Constructing a Randomized Strategy

In this section we show how a randomized strategy may be constructed given a finite length, fixed strategy so that a lower worst-case cost ratio is obtained under a large class of cost functions. In the process of doing this we also reveal how randomized strategies perform better for this criterion. We will limit our discussion to discrete strategies $\mathbf{u} \in U$, and hence assume that the object location is integer-valued. The cost function $C(x)$ thus only needs to be defined at integers values of x. Subsequently, for the remainder of this section whenever we write $1 \leq x \leq L$ it is implied that x only takes integer values between 1 and L.

Definition 7: An increasing cost function $C(x)$ belongs to the class $\mathbb{C}^*$ if for all integers $3 \leq x \leq L$,

$$C(x) < C(x-1) + \frac{C(x-1)^2}{\sum_{i=1}^{x-2} C(i)} . \tag{11}$$

Note that both the linear and quadratic cost functions given earlier satisfy (11). This constraint essentially limits the amount of increase $C(x) - C(x-1)$ in the cost function.

This definition is introduced for technical reasons for proving Theorem 1 below on a specific construction of a randomized strategy, as this constraint simplifies the identification of worst-case locations. It should be noted that it is possible to apply similar methods under more general cost functions by considering a different set of possible worst-case locations.

Consider any nonrandom TTL sequence given by the finite length vector $\mathbf{u} = [u_1, u_2, \cdots, u_N]$, where $u_N = L$ and $u_1 < u_2 < ... < u_{N-1} < u_N$. Also define $u_0 = 0$. We have the following result.

Lemma 3: Let an integer x^*, $1 \le x^* \le L$, be such that

$$\rho^{\mathbf{u}} = \frac{J^{\mathbf{u}}_{x^*}}{C(x^*)} = \max_{1 \le x \le L} \frac{J^{\mathbf{u}}_x}{C(x)} . \tag{12}$$

If the cost $C(x) \in \mathbb{C}^*$, then either $x^* = u_n + 1 < u_{n+1}$ for some $0 \le n \le N-1$ or $x^* = u_N = L$.

What this lemma says is that if $C(\cdot) \in \mathbb{C}^*$, then the worst-case location is either immediately following one of the TTL values in the sequence $\mathbf{u}$, or at the boundary L for any given deterministic TTL sequence $\mathbf{u} = [u_1, u_2, \cdots, u_N]$. This result is intuitively clear in that the worst location is the closest point that is outside some searched area u_n. For such a TTL sequence, define a set S as follows:

$$S = \left\{ 1 \le x^* \le L,\ x^* \in \mathbb{Z}^+ : \frac{J^{\mathbf{u}}_{x^*}}{C(x^*)} = \max_{1 \le x \le L} \frac{J^{\mathbf{u}}_x}{C(x)} \right\} , \tag{13}$$

which is essentially the set of all the worst-case location values. Let $|S|$ denote the number of such points at which the maximum is achieved. It follows from Lemma 3 that $1 \le |S| \le N$.

We now construct a randomized strategy $\hat{\mathbf{u}}$ from the fixed sequence $\mathbf{u} = [u_1, u_2, \cdots, u_N]$ as follows, referred to as Construction ($\mathbf{T}$).

(T.1) For all u_m such that $u_m+1 \notin S$ and $1 \le m \le N-1$, set $\tilde{u}_m = u_m$. For any u_m such that $u_m+1 \in S$ and $1 \le m \le N-1$, set $\tilde{u}_m = u_m + 1$. In addition, set $\tilde{u}_N = u_N = L$ regardless of whether it belongs to set S.

(T.2) If $u_N \notin S$ and $1 \notin S$, then define $\tilde{\mathbf{u}} = [\tilde{u}_1, \tilde{u}_2, ...\tilde{u}_N]$.

(T.3) If $u_N \notin S$ and $1 \in S$, define $\tilde{\mathbf{u}} = [1, \tilde{u}_1, \tilde{u}_2, ...\tilde{u}_N]$, noting that it follows from Lemma 3 that $\tilde{u}_1 \ge u_1 > 1$. Also note that $\tilde{\mathbf{u}}$ defined in this case differs from that in (T.2) in the insertion of a "1" before $\tilde{u}_1$.

(T.4) If $u_N \in S$ and $1 \in S$, define $\tilde{\mathbf{u}} = [1, \tilde{u}_1, \tilde{u}_2, ...\tilde{u}_{N-2}, \tilde{u}_N]$. Note that $\tilde{\mathbf{u}}$ defined in this case differs from that in (T.3) in the removal of $\tilde{u}_{N-1}$.

(T.5) Finally, if $u_N \in S$ and $1 \notin S$, then define $\tilde{\mathbf{u}} = [\tilde{u}_1, \tilde{u}_2, ...\tilde{u}_{N-2}, \tilde{u}_N]$. Note that $\tilde{\mathbf{u}}$ defined in this case differs from that in (T.4) in the removal of the "1".

(T.6) Now construct the random TTL sequence $\hat{\mathbf{u}} = [\hat{u}_1, \hat{u}_2, ..., \hat{u_M}]$ by randomly selecting one of two sequences: with probability $1-p$ we employ the original TTL sequence $\mathbf{g}$, and with probability p we employ the modified TTL sequence $\tilde{\mathbf{u}}$ defined in (T.2)-(T.5) for each of the four possible cases, respectively, where p is given by

$$0 < p < \min\left\{1, \min_{x \notin S}\left\{ \left(\rho^{\mathbf{u}} - \frac{J^{\mathbf{u}}_x}{C(x)}\right) \frac{C(x)}{M_L} \right\}\right\} , \tag{14}$$

and

$$M_j = \sum_{i \in S, i \le j} \left(C(i) - C(i-1)\right), \quad 1 \le j \le L ,$$

where $C(0) = 0$.

Theorem 1: Consider any nonrandom TTL sequence given by the integer-valued finite length vector $\mathbf{u} = [u_1, u_2, ...u_N]$, where $u_N = L$ and $u_1 < u_2 < ... < u_{N-1} < u_N$. Generate a new random TTL sequence $\hat{\mathbf{u}}$ using Construction ($\mathbf{T}$). If the cost function belongs to the class $\mathbb{C}^*$, then:

$$\rho^{\hat{\mathbf{u}}} = \max_{1 \le x \le L} \frac{J^{\hat{\mathbf{u}}}_x}{C(x)} < \max_{1 \le x \le L} \frac{J^{\mathbf{u}}_x}{C(x)} = \rho^{\mathbf{u}} . \tag{15}$$

Therefore there exists at least one random TTL sequence given by $\hat{\mathbf{u}}$ that achieves a lower worst-case cost than that using the nonrandom sequence $\mathbf{u}$.

The proof of this theorem is given in the Appendix. The key observation here is that under this construction, for any $x \notin S$ we have $\frac{J^{\hat{\mathbf{u}}}_x}{C(x)} - \frac{J^{\mathbf{u}}_x}{C(x)} > 0$, and $\frac{J^{\hat{\mathbf{u}}}_x}{C(x)} < \rho^{\mathbf{u}}$. What this means is that randomizing some of the TTL values of $\mathbf{u}$ increases the cost ratio at points $x \notin S$ but not sufficient to exceed the worst-case cost ratio $\rho^{\mathbf{u}}$. On the other hand, we also have that for $x \in S$, $\frac{J_x^{\hat{\mathbf{u}}}}{C(x)} < \frac{J_x^{\mathbf{u}}}{C(x)}$, which means that the cost ratio is decreased at all x for which the worst-case cost is achieved under $\mathbf{u}$. Combining these observations, the overall effect of this particular randomization construction is to decrease the value of $\rho^{\mathbf{u}}$ by lowering the cost ratio at $x \in S$ at the expense of increasing the cost ratio at $x \notin S$. This is formally shown in the proof. In essence, this randomization attempts to "spread" the cost at the worst-cast points to their neighboring points in order to bring down the worst-case cost. This is also the fundamental intuition behind why randomization performs better for the worst-case critetion.

The construction here is for the class $\mathbb{C}^*$ cost functions. It is however conceivable that similar randomizations can be constructed for more general cost functions, by considering the corresponding worst-case points resulting from the cost functions.

V. 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 an infinitely large network and TTL sequences of infinite length that satisfy Definition 5. 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. It will become evident that asymptotically optimal TTL sequences can also perform very well in a network of arbitrary finite dimension. In particular, as will be shown, under the derived infinite-length strategies the worst-case cost ratio is reached asymptotically from below as the object location reaches ∞, and hence the cost ratio at any finite object location is less than the worst-case cost ratio.

In what follows we will first derive a tight lower bound on the worst-case cost ratio for both continuous and discrete strategies. We then introduce a particular randomized continuous strategy and show that this strategy has a worst-case cost ratio matching the lower bound, therefore proving that this strategy is optimal in the worst-case. We then derive a discrete strategy from the optimal continous strategy and show that it also achieves the lower bound.

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

In deriving a tight lower bound on the worst-case cost ratio, we first use Yao's minimax principle [10] and Lemma 2 to obtain the following inequality:

Lemma 4:

$$\sup_{\{p_X(x)\}} \inf_{\mathbf{u}\in U^d} \frac{J_X^{\mathbf{u}}}{E[C(X)]} \leq \inf_{\mathbf{u}\in U} \sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} . \tag{16}$$

Proof: First note that for any given object probability distribution, there exists an optimal strategy that is deterministic. Hence the following holds:

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

We also have the following in interchanging the supremum and infimum, see for example [11]:

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

Finally, applying Lemma 2 to the above equality and inequality establishes (16). ■

The corresponding continuous version of Lemma 4 is straightforward.

Lemma 5:

$$\sup_{\{f_X(x)\}} \inf_{\mathbf{v}\in V^d} \frac{J_X^{\mathbf{v}}}{E[C(X)]} \leq \inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} . \tag{17}$$

The proof of this lemma is similar to the discrete case and is not repeated.

We now use the above results to first derive a lower bound on the minimum worst-case cost ratio under continuous strategies. Using (17), 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 $C(x) \in \mathbb{C}$.

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

[1] A special case of this distribution where cost $C(\cdot)$ is linear, also known as the Zipf distribution, was studied in [8] for which the optimal deterministic strategy was computed. Here we essentially follow the same method (generalized to any cost function in $\mathbb{C}$) to derive the class of optimal strategies.

where $v_0 = 1$ is assumed for simplicity of notation:

$$J_X^{\mathbf{v}} = \sum_{j=1}^{\infty} C(v_j)\bar{F}(v_{j-1}) = \sum_{j=1}^{\infty} C(v_j) \left(\frac{C(v_{j-1})}{C(1)} \right)^{-\alpha} . \tag{18}$$

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

$$\frac{\partial J_X^{\mathbf{v}}}{\partial v_j} = [C(v_{j-1})^{-\alpha} - \alpha C(v_{j+1}) C(v_j)^{-\alpha-1}] \frac{\partial C(v_j)}{\partial v_j} (C(1))^{\alpha} = 0 , \tag{19}$$

for all $j \geq 1$. Since both the derivative of the cost function and $C(1)$ are strictly positive, the term enclosed in brackets in (19) must be 0. Hence for a given fixed v_1, the optimal strategy is to recursively choose v_j that satisfies the following equation for all $j \geq 1$:

$$C(v_{j+1}) = \frac{C(v_j)}{\alpha} \left(\frac{C(v_j)}{C(v_{j-1})} \right)^{\alpha} . \tag{20}$$

Note that this optimal sequence satisfies the following:

$$\bar{F}(v_j) C(v_{j+1}) = \frac{C(v_{j+1}) C(1)^{\alpha}}{C(v_j)^{\alpha}} = \left(\frac{C(v_j)}{C(v_{j-1})} \right)^{\alpha} \frac{C(v_j) C(1)^{\alpha}}{\alpha C(v_j)^{\alpha}} = \frac{C(1)^{\alpha}}{C(v_{j-1})^{\alpha}} \frac{C(v_j)}{\alpha} = \bar{F}(v_{j-1}) \frac{C(v_j)}{\alpha} . \tag{21}$$

Summing both sides of (21) from $j = 1$ to $j = \infty$ and multiplying by α gives:

$$\alpha \sum_{j=1}^{\infty} \bar{F}(v_j) C(v_{j+1}) = \sum_{j=1}^{\infty} \bar{F}(v_{j-1}) C(v_j) \implies \alpha \left(\sum_{i=0}^{\infty} \bar{F}(v_j) C(v_{j+1}) - C(v_1) \right) = \sum_{j=1}^{\infty} \bar{F}(v_{j-1}) C(v_j) . \tag{22}$$

Substituting this in the definition of $J_X^{\mathbf{v}}$ gives:

$$\alpha J_X^{\mathbf{v}} - \alpha C(v_1) = J_X^{\mathbf{v}} \implies J_X^{\mathbf{v}} \frac{\alpha - 1}{\alpha} = C(v_1) . \tag{23}$$

On the other hand, the mean of the object location can be calculated as follows, noting that X takes values on $[1, \infty)$:

$$\begin{aligned} E\left[C(X)\right] &= \int_0^{\infty} Pr\left(C\left(X\right) > x\right) \, dx = C(1) + \int_{C(1)}^{\infty} \bar{F}(C^{-1}(x)) \, dx \\ &= C(1) + \int_{C(1)}^{\infty} \left[\frac{C(C^{-1}(x))}{C(1)} \right]^{-\alpha} \, dx = C(1) + \frac{1}{C(1)^{-\alpha}} \int_{C(1)}^{\infty} x^{-\alpha} \, dx = \frac{\alpha}{\alpha - 1} C(1) . \end{aligned} \tag{24}$$

(23) and (24) imply that for a sequence defined by a given v_1 and using the recursion given by (20), the cost ratio is

$$\frac{J_X^{\mathbf{v}}}{E\left[C\left(X\right)\right]} = J_X^{\mathbf{v}} \frac{(\alpha - 1)}{\alpha C(1)} = \frac{C(v_1)}{C(1)} . \tag{25}$$

This result implies that for a given α, the sequence that generates the smallest cost ratio will follow the recursion (20) and use the smallest possible value of v_1. However, not all values of v_1 lead to an increasing sequence **v**, which is obviously a requirement for an optimal strategy. In fact, we have the following result:

Lemma 6: Consider any infinite length sequence $\mathbf{v} = [v_1, v_2, ..]$, where v_1 is some positive constant and v_k for $k \geq 2$ is generated by the recursion given by (20). Then **v** is an increasing sequence if and only if the following condition holds:

$$\frac{C(v_1)}{C(1)} \geq \alpha^{(\sum_{k=1}^{\infty} \alpha^{-k})} = \alpha^{\frac{1}{\alpha-1}} . \tag{26}$$

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

Lemma 7: For any $C(x) \in \mathbb{C}$, the worst-case cost ratio of any continuous strategy is lower-bounded by e, i.e.:

$$\inf_{\mathbf{v} \in V} \sup_{x \in [1, \infty)} \frac{J_x^{\mathbf{v}}}{C(x)} \geq e . \tag{27}$$

This result implies that if we can obtain a TTL sequence whose worst-case ratio is e, then this 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 8: Assume that the cost function $C(x) \in \mathbb{C}$. 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 $\left[1, C^{-1}(r \cdot C(1))\right)$, 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 $F_{v_1}\left(C^{-1}(r \cdot C(1))\right) = 1$.

(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.

Lemma 8: Consider any strategy $\mathbf{v}[r, F_{v_1}(x)]$ constructed using steps (J.1) and (J.2) above. Assume $C(x) \in \mathbb{C}$. Let $\bar{F}_{v_1}(y) = 1 - F_{v_1}(y)$. Then the worst-case cost ratio is given by:

$$\sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} = \sup_{1 \leq z < r} \left\{ \frac{r}{r-1} \frac{h(r) + (r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} \right\} , \tag{28}$$

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

$$h(z) = C(1) + \int_{C(1)}^{z \cdot C(1)} \bar{F}_{v_1}(C^{-1}(y))\, dy . \tag{29}$$

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:

$$\Phi(z) = \frac{r}{r-1} \frac{h(r) + (r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} . \tag{30}$$

From Lemma 8, the worst-case cost ratio of $\mathbf{v}$ is the supremum of $\Phi(z)$ over the range $1 \leq z < r$. Note that the following four boundary conditions are true for any function $h(z)$ as defined by (29):

$$h(1) = C(1)\ ,\ h(r) = E[C(v_1)]\ ,\ h'(1) = C(1)\ ,\ h'(r) = 0\ . \tag{31}$$

Therefore we have:

$$\Phi(1) = \frac{r}{r-1} \frac{[h(r) + (r-1)h(1)]}{C(1)} - r\frac{h'(1)}{C(1)} = \frac{r}{r-1} \frac{h(r)}{C(1)} , \tag{32}$$

$$\Phi(r) = \frac{r}{r-1} \frac{h(r) + (r-1)h(r)}{rC(1)} - r\frac{h'(r)}{C(1)} = \frac{r}{r-1} \frac{h(r)}{C(1)} . \tag{33}$$

Theorem 2: Assume $C(x) \in \mathbb{C}$. We have

$$\inf_{\mathbf{v} \in V} \sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} = e . \tag{34}$$

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 $1 \leq x < C^{-1}(eC(1))$, and $v_k^* = C^{-1}\left(e^{k-1} C\left(v_1^*\right)\right)$ for all positive integers k.

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

$$\begin{aligned} h(z) &= C(1) + \int_{C(1)}^{zC(1)} \bar{F}_{v_1}\left(C^{-1}(y)\right)\, dy = C(1) + \int_{C(1)}^{z \cdot C(1)} \left(1 - ln\ \frac{y}{C(1)}\right)\, dy \\ &= C(1)\left[z - z(ln\ z - 1) - 1\right] = C(1)\left[2z - z\ ln\ z - 1\right]\ . \end{aligned} \tag{35}$$

Note that $h(e) = C(1)\,(e-1)$ and $h'(z) = C(1)\,[1 - ln\ z]$. Therefore we can calculate $\Phi(z)$ for $1 \leq z < r$ as follows:

$$\Phi(z) = \frac{e}{e-1} \frac{C(1)(e-1) + (e-1)C(1)\left[2z - zln\ z - 1\right]}{zC(1)} - \frac{eC(1)(1 - ln\ z)}{C(1)} = e\ . \tag{36}$$

Hence it is clear from Lemma 8 that the worst-case cost ratio of this sequence is e. Combine this with Lemma 7 which showed the worst-case cost ratio of any continuous strategy is lower bounded by e, we complete the proof. ■

In what follows we illustrate the idea behind the derivation of this optimal strategy. As mentioned in Section IV, one way to randomize or improve a strategy is to decrease the cost ratio at the worst location points by spreading the cost to neighboring points. This leads one to conjecture that under an optimal strategy, the cost ratio as a function of object location x does not have any local maxima or minima, producing a smooth curve. Led by this conjecture, we set out to find the type of cdf $\bar{F}_{v_1}(y)$ and the corresponding $h(z)$ that will produce a flat cost ratio curve, i.e. $\Phi(z) = \Phi(1) = \Phi(r) = \frac{r}{r-1}\frac{h(r)}{C(1)}$ for all $1 \leq z < r$, using Lemma 8.

This flat cost ratio curve can be achieved if and only if $h(z)$ satisfies the following differential equation:

$$\frac{r}{r-1}\frac{h(r)+(r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} = \frac{r}{r-1}\frac{h(r)}{C(1)} \implies -z\cdot h'(z) - \frac{z-1}{r-1}h(r) + h(z) = 0\ , \tag{37}$$

for all $1 \leq z < r$. The equation on the righthand-side of (37) is a first-order linear differential equation and it can be shown that the general solution satisfies the following for all $1 \leq z < r$:

$$h(z) = z\cdot c + \frac{z}{r-1}\left(-\frac{h(r)}{z} - h(r)\cdot ln(z)\right) = z\cdot c - \frac{h(r)}{r-1} - \frac{h(r)}{r-1}z\cdot ln(z)\ , \tag{38}$$

where the values of the constant c and $h(r)$ can be determined by the boundary conditions in (31) as follows. We have:

$$h'(z) = c - \frac{h(r)}{r-1}ln(z) - \frac{h(r)}{r-1}\ . \tag{39}$$

Therefore

$$C(1) = h'(1) = c - \frac{h(r)}{r-1}ln(1) - \frac{h(r)}{r-1} = c - \frac{h(r)}{r-1} \implies c = C(1) + \frac{h(r)}{r-1} \tag{40}$$

$$0 = h'(r) = c - \frac{h(r)}{r-1}ln(r) - \frac{h(r)}{r-1} \tag{41}$$

Substituting (40) into (41), we obtain:

$$0 = C(1) - \frac{h(r)}{r-1}ln(r) \implies h(r) = C(1)\frac{r-1}{ln(r)}\ , \tag{42}$$

which means that $c = C(1)\left[1 + \frac{1}{ln(r)}\right]$. Hence (38) becomes:

$$h(z) = C(1)\left[z\left(1+\frac{1}{ln(r)}\right) - \frac{1}{ln(r)} - \frac{z\cdot ln(z)}{ln(r)}\right] = C(1)\frac{z\,[ln(r)+1-ln(z)]-1}{ln(r)}\ . \tag{43}$$

Differentiating this result gives us the tail distribution and corresponding cdf, for $1 \leq z < r$, and $1 \leq y < C^{-1}(rC(1))$:

$$\begin{aligned} &\bar{F}_{v_1}\left(C^{-1}(zC(1))\right)C(1) = h'(z) = C(1)\left[1 - \frac{ln(z)}{ln(r)}\right] \\ \implies\ & F_{v_1}\left(C^{-1}(zC(1))\right) = \frac{ln(z)}{ln(r)} \\ \implies\ & F_{v_1}(y) = \frac{1}{\ln r}\ln\frac{C(y)}{C(1)}\ . \end{aligned} \tag{44}$$

Equation (44) gives the family of cdf functions that result in a flat cost ratio curve for a give r. Finally, the corresponding worst-case cost ratio is given by substituting (42) into (32):

$$\sup_{x\in[1,\infty)}\frac{J_x^{\mathbf{u}}}{C(x)} = \frac{r}{r-1}\frac{h(r)}{C(1)} = \frac{r}{ln(r)}\ . \tag{45}$$

By differentiating and noting convexity, we find that using $r = e$ obtains a worst-case cost ratio of e. Hence, this must be an optimal worst-case strategy following the same argument given in the proof of Theorem 2.

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 defined as follows. The first TTL value is a random variable v_1^* with cdf $F_{v_1^*}(z) = ln\, z$ for $1 \leq z < e$. Successive TTL values are defined as $v_k^* = e^{k-1}v_1^*$.

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 $C(\cdot)$. We therefore in this section limit our attention to a subclass of $\mathbb{C}$ and derive optimal strategies for this subclass, given by the following definition:

Definition 9: A function $C(\cdot) \in \mathbb{C}$ belongs to the class $\mathbb{C}_q$ for some $q \geq 1$ if: (1) $\lim_{x\to\infty} \frac{C(x+1)}{C(x)} = q$ and (2) $\frac{C(x+1)}{C(x)} \geq q$ for all $x \in [1, \infty)$.

Note that since $C(\cdot)$ is strictly increasing, for $q = 1$ condition (2) is automatically satisfied. The case of $q = 1$ also contains all polynomial cost functions. The case of $q > 1$ includes for example exponential cost functions of the form $q^{\,x}$. Therefore even though this is a subclass of $\mathbb{C}$, it remains very general.

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 $Pr(X_\alpha > x) = \left(\frac{C(x)}{C(1)}\right)^{-\alpha}$ for all $x \geq 1$ and some constant $\alpha > 1$. We have the following result:

Lemma 9: If $C(\cdot) \in \mathbb{C}_q$, then:

$$\lim_{\alpha\to 1^+} \frac{E[C(X_\alpha + 1)]}{E[C(X_\alpha)]} = q \,. \tag{46}$$

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

The lower-bound on the best worst-case cost ratio is given by the following lemma.

Lemma 10: Suppose $C(\cdot) \in \mathbb{C}_q$. We have:

$$\inf_{\mathbf{u}\in U} \sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} \geq \frac{e}{q} \,. \tag{47}$$

Proof: Consider any $\mathbf{u} \in U$. For any integer $x \geq 2$:

$$\frac{J_x^{\mathbf{u}}}{C(x)} = \lim_{\epsilon\to 0} \frac{J_{x-1+\epsilon}^{\mathbf{u}}}{C(x+\epsilon)} = \sup_{y\in[x-1,x)} \frac{J_y^{\mathbf{u}}}{C(y+1)} \,, \tag{48}$$

since $J_{x-1+\epsilon}^{\mathbf{u}} = J_x^{\mathbf{u}}$ for all $0 < \epsilon \leq 1$, and $C(\cdot)$ is strictly increasing. Hence we have:

$$\sup_{x\in\mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} = \sup\left\{\frac{J_1^{\mathbf{u}}}{C(1)}, \sup_{x\in\mathbb{Z}^+,\, x\geq 2} \sup_{y\in[x-1,x)} \frac{J_y^{\mathbf{u}}}{C(y+1)}\right\} = \sup\left\{\frac{J_1^{\mathbf{u}}}{C(1)}, \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{u}}}{C(x+1)}\right\} \tag{49}$$

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:

$$\sup_{\{f_X(X)\}} \frac{J_X^{\mathbf{v}}}{E[C(X+1)]} = \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x+1)} \,, \tag{50}$$

Thus similarly to Lemma 5, we have:

$$\sup_{\{f_X(x)\}} \inf_{\mathbf{v}\in V^d} \frac{J_X^{\mathbf{v}}}{E[C(X+1)]} \leq \inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x+1)} \,. \tag{51}$$

Again, any lower bound can be found by first selecting a location distribution $f_X(x)$ and deriving the optimal deterministic strategy.

Consider the tail distribution given by $Pr(X_\alpha > x) = \left(\frac{C(x)}{C(1)}\right)^{-\alpha}$ for all $x \geq 1$ and some constant $\alpha > 1$. It was shown earlier in Section V-A that for object location X_α, the optimal cost ratio is $\alpha^{\frac{1}{\alpha-1}}$. This approaches e as $\alpha \to 1^+$. Hence using Lemma 9 we have:

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

Using this result in (51) and $U \subset V$ gives us:

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

where the last inequality follows from (49). ∎

This result says that if we can find a discrete strategy whose worst-case cost ratio is e/q for some $C(\cdot) \in \mathbb{C}_q$, then this strategy must be optimal among all admissible discrete strategies. Unfortunately it appears difficult to find strategies matching this lower bound for all $C(\cdot) \in \mathbb{C}_q$. It is, however, possible to do so for the special case of $q = 1$, which includes all polynomials as mentioned earlier.

We start by upper bounding the best worst-case cost ratio of discrete strategies for any $C(\cdot) \in \mathbb{C}$. For any continuous strategy $\mathbf{v}^* = [v_1^*, v_2^*, \cdots]$, we use $\mathbf{u}^* = \lfloor \mathbf{v}^* \rfloor$ to denote the discrete strategy $\mathbf{u}^* = [u_1^*, u_2^*, ...]$ obtained by setting $u_k^* = \lfloor v_k^* \rfloor$ for all k.

Lemma 11: Suppose $C(\cdot) \in \mathbb{C}$. We have

$$\inf_{\mathbf{u} \in U} \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} \leq e \, . \tag{52}$$

Proof: Let $\mathbf{v}^*$ denote the strategy $\mathbf{v}^*[e, \ln \frac{C(x)}{C(1)}]$ and $\mathbf{u}^* = \lfloor \mathbf{v}^* \rfloor$. For any positive integer k, u_k^* takes integer values between $\lfloor C^{-1}\left(e^{k-1}C(1)\right) \rfloor$ and $\lfloor C^{-1}\left(e^{k}C(1)\right) \rfloor$. In addition, $\lfloor C^{-1}\left(e^{k}C(1)\right) \rfloor$ is a nondecreasing sequence with respect to integer values of k, and approaches ∞ as k approaches ∞. Fix the object location as a positive integer x, and choose the smallest integer k such that $x \leq \lfloor C^{-1}\left(e^{k}C(1)\right) \rfloor$. Note that $E\left[C(u_k^*)\right] = E\left[C\left(\lfloor v_k^* \rfloor\right)\right] \leq E\left[C\left(v_k^*\right)\right]$ for all integers k. Since x is a positive integer, we have $Pr\left(u_k^* < x\right) = Pr\left(\lfloor v_k^* \rfloor < x\right) = Pr\left(v_k^* < x\right)$. Hence we have the following for this x:

$$\begin{aligned} J_x^{\mathbf{u}^*} &= \sum_{j=1}^{k} E[C\left(u_j^*\right)] + Pr(u_k^* < x) E[C\left(u_{k+1}^*\right) | u_k^* < x] \leq \sum_{j=1}^{k} E\left[C\left(v_j^*\right)\right] + Pr\left(\lfloor v_k^* \rfloor < x\right) E\left[C\left(v_{k+1}^*\right) | \lfloor v_k^* \rfloor < x\right] \\ &= J_x^{\mathbf{v}^*} \leq eC(x) \, , \end{aligned}$$

where the last inequality holds because the worst-case cost ratio for $\mathbf{v}^*$ is e as proven in Theorem 2. Since this result holds for all integers x, we have $\sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}^*}}{C(x)} \leq e$. ■

Note that when $q = 1$, the lower and upper bounds in Lemmas 10 and 11 match. Therefore:

Theorem 3: Suppose $C(\cdot) \in \mathbb{C}_1$. We have

$$\inf_{\mathbf{u} \in U} \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{C(x)} = e \, . \tag{53}$$

Moreover, this worst-case cost ratio is obtainable by strategy $\mathbf{u}^* = \lfloor \mathbf{v}^* \rfloor$, where $\mathbf{v}^*$ denotes the strategy $\mathbf{v}^*[e, \ln \frac{C(x)}{C(1)}]$.

Since $\mathbb{C}_1$ includes all increasing polynomial functions, the optimal strategy given in Theorem 3 can be used when cost is given by or can be approximated by a polynomial function, which is not a very restrictive assumption.

VI. Performance Comparison and Discussion

In this section we first compare the performance of the optimal randomized strategy with that of a 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 2 we compare the cost ratio of the optimal discrete strategy given by Theorem 3 to that of a non-random TTL sequence given by the California Split search $u_k = 2^{k-1}$ for all k under the linear cost function $C(k) = k$. $u_k = \lfloor e^{k-1} \rfloor$ for all 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. The construction **(T)** we used in Section IV, and the way in which we derived the optimal continuous strategy in Section V-C are essentially both attempting to achieve this effect. 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.

B. Other Performance Measures

Next we discuss alternative performance measures for analyzing randomized search strategies. We will again assume that $C(x) \in \mathbb{C}$, 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 $\mathbf{v}$, as the same randomized strategy can result in

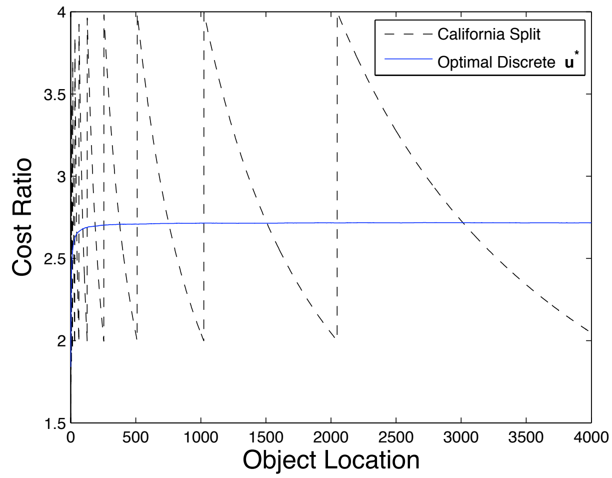

Fig. 2. Cost ratio as a function of object location for the optimal discrete sequence $\mathbf{u}^*$ described in Theorem 3, and California Split search defined by $u_k = 2^{k-1}$ for all k. Cost is assumed to be linear.

different realizations. This leads us to consider the competitive ratio with respect to an *adaptive offline adversary* [10] 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,

$$\Gamma_X^{\mathbf{v}} = \sup_{\tilde{\mathbf{v}} \in \Upsilon^{\mathbf{v}}} \frac{J_X^{\tilde{\mathbf{v}}}}{E\left[C(X)\right]} \,, \tag{54}$$

where $\Upsilon^{\mathbf{v}}$ denotes the set of all possible realizations of strategy $\mathbf{v}$. Let the *worst-case worst realization cost ratio* $\Gamma^{\mathbf{v}}$ denote the maximum of $\Gamma_X^{\mathbf{v}}$ over all possible object locations. Then the performance of a search strategy against an adaptive offline adversary can be measured by the following competitive ratio (worst-case, worst-realization):

$$\Gamma^{\mathbf{v}} = \sup_{\{f_X(x)\}} \Gamma_X^{\mathbf{v}} = \sup_{x \in [1,\infty)} \Gamma_x^{\mathbf{v}} \,, \tag{55}$$

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

As discussed in [10], 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. On the other hand, the minimum worst-case average cost ratio of all deterministic real-valued strategies under $C(x) \in \mathbb{C}$ can be shown to be 4 [2]. Therefore, we have the following:

$$\inf_{\mathbf{v} \in V} \Gamma^{\mathbf{v}} = \inf_{\mathbf{v} \in V^d} \sup_{\{f_X(x)\}} \frac{J_X^{\mathbf{v}}}{E[C(X)]} = 4 \,. \tag{56}$$

Similarly, let $\gamma_X^{\mathbf{v}}$ and $\gamma_x^{\mathbf{v}}$ denote the *best-realization cost ratio* for 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 $\Lambda_x^{\mathbf{v}}$ denote the variance of the search cost when using strategy $\mathbf{v}$ and fixed object location $x \in [1,\infty)$. Therefore, $\Lambda_x^{\mathbf{v}}/C(x)^2$ is the variance of the ratio $j^{\mathbf{v}}/C(x)$ when object location is x.

We now examine these quantities for the class of jointly defined continuous strategies $\mathbf{v}[r, F_{v_1}(x)]$ as given by Definition 8. We begin with the following theorem:

Theorem 4: Consider a real-valued randomized strategy $\mathbf{v}[r, F_{v_1}(x)]$ that is constructed as given by Definition 8, and a real-valued deterministic strategy $\hat{\mathbf{v}} = [\hat{v}_1, \hat{v}_2, ...]$ defined by $\hat{v}_k = C^{-1}\left(r^{k-1}C(1)\right)$ for all integers $k \geq 1$. Then we have:

$$\Gamma^{\mathbf{v}} \leq \rho^{\hat{\mathbf{v}}} = \frac{r^2}{r-1} \,. \tag{57}$$

[2]This can be shown in a similar manner to that used in [8] for discrete strategies under linear cost. In particular, in Section VII we establish an equivalency between linear and general cost functions, which can used to show that 4 is minimum worst-case cost ratio among deterministic strategies.

In other words, the worst-case worst-realization cost ratio for the random strategy $\mathbf{v}$ is upper bounded by the worst-case average cost ratio for the deterministic sequence $\hat{\mathbf{v}}$.

Proof: We first show that $\rho^{\hat{\mathbf{v}}} = \frac{r^2}{r-1}$. Fix any finite $x \geq 1$. There must exist a k such that $C^{-1}\left(r^{k-1}C(1)\right) \leq x < C^{-1}\left(r^{k}C(1)\right)$. Let $x_k = C^{-1}\left(r^{k}C(1)\right)$ for all integers k. Then the search cost can be calculated as follows:

$$J_x^{\hat{\mathbf{v}}} = \sum_{j=1}^{k+1} C\left(\hat{v}_j\right) = \sum_{j=1}^{k+1} r^{j-1}C(1) = C(1)\frac{r^{k+1}-1}{r-1} \; . \tag{58}$$

Since this holds for all x between x_{k-1} and x_k we have:

$$\sup_{x_{k-1}<x\leq x_k} \frac{J_x^{\hat{\mathbf{v}}}}{C(x)} = \lim_{x\to x_{k-1}^{+}} \frac{(r^{k+1}-1)C(1)}{(r-1)C(x)} = \frac{r^2-r^{-k+1}}{r-1} \; . \tag{59}$$

Hence we have:

$$\rho^{\hat{\mathbf{v}}} = \sup_{x\in[1\infty)} \frac{J_x^{\hat{\mathbf{v}}}}{C(x)} = \sup_{k\in\mathbb{Z}^+} \left\{ \sup_{x_{k-1}<x\leq x_k} \frac{J_x^{\hat{\mathbf{v}}}}{C(x)} \right\} = \sup_{k\in\mathbb{Z}^+} \frac{r^2-r^{-k+1}}{r-1} = \lim_{k\to\infty} \frac{r^2-r^{-k+1}}{r-1} = \frac{r^2}{r-1} \; . \tag{60}$$

We next show that $\Gamma^{\mathbf{v}} \leq \frac{r^2}{r-1}$. Fix the object location x. Again, there must exist exactly one k such that $x_{k-1} \leq x < x_k$. 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 $\mathbf{v}$.

It can be shown that 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.

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:

$$\Gamma_x^{\mathbf{v}} \leq \lim_{\tilde{v}_1\to\left(C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right)^{-}} \left\{ \frac{1}{C(x)} \sum_{j=1}^{k+1} r^{j-1}C\left(\tilde{v}_1\right) \right\} = \frac{r^{k+1}-1}{(r-1)r^{k-1}} = \frac{r^2-r^{-k+1}}{r-1} \; . \tag{61}$$

It easily follows that:

$$\Gamma^{\mathbf{v}} \leq \lim_{x\to\infty} \Gamma_x^{\mathbf{v}} \leq \lim_{k\to\infty} \frac{r^2-r^{-k+1}}{r-1} = \frac{r^2}{r-1} \; , \tag{62}$$

thus completing the proof. ■

The inequality in this theorem 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 (61) and (62) become equalities. This is true when $F_{v_1}(x) = (\ln \frac{C(x)}{C(1)})/\ln\ r$, and hence strategies with this family of cdf have worst-case worst-realization cost ratio value of $r^2/(r-1)$. Therefore the worst-realization can be adjusted by selecting the appropriate value of r.

Similarly, the best-realization cost ratio of these types of strategies can be calculated for object location x, where we have $C^{-1}\left(r^{k-1}C(1)\right) \leq x < C^{-1}\left(r^{k}C(1)\right)$ for some positive integer k. 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 becomes:

$$\gamma_x^{\mathbf{v}} = \frac{\sum_{j=1}^{k} r^{j-1}\frac{C(x)}{r^{k-1}}}{C(x)} = \frac{r-r^{-k+1}}{r-1} \; . \tag{63}$$

Clearly, the asymptotic (i.e. as x approaches ∞) best realization is $\frac{r}{r-1}$.

The variance of this strategy is calculated in Appendix G. The key result from these calculations is that the asymptotic limit of the variance (with respect to object location) is given by:

$$\lim_{x\to\infty} \frac{\Lambda_x^{\mathbf{v}}}{C(x)^2} = \frac{r^4-r^2}{2(ln\, r)(r-1)^2} - \frac{r^2}{(ln\, r)^2} \; . \tag{64}$$

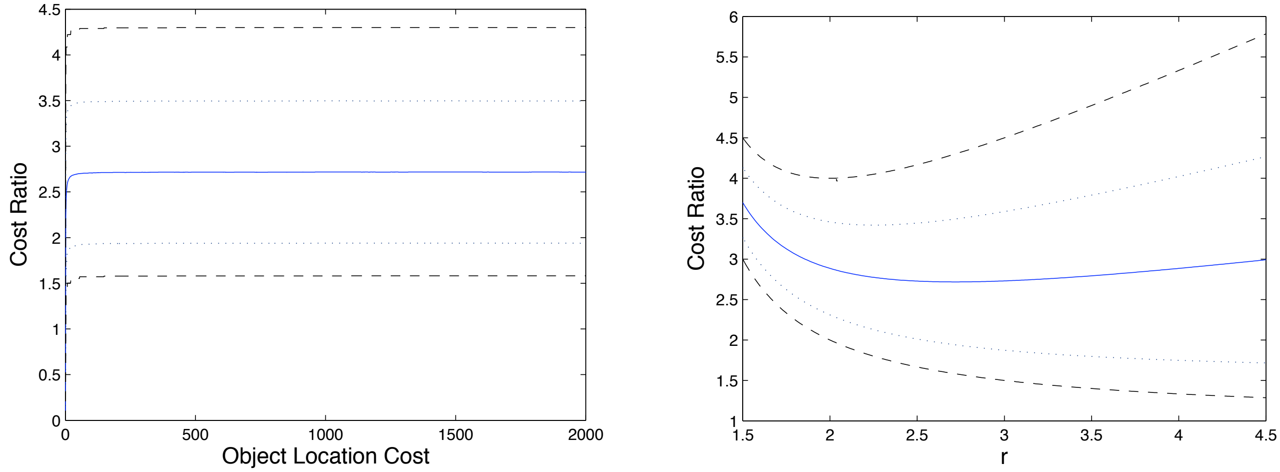

Fig. 3. (LEFT): Performance of optimal continuous strategy (presented in Theorem 2) as a function of object location cost. Worst and best realization cost ratio (top and bottom dashed lines), average cost ratio (solid), and average cost ratio $+/-$ one standard deviation (dotted) are shown. (RIGHT): Performance of continuous strategies given by Definition 8 and using the cdf given by (44), 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.

The cost ratio under the linear cost function for the optimal continuous strategy is depicted in Figure 3 (LEFT) as a function of object location. Note that under these metrics, this strategy's performance does not change much with respect to object location.

For strategies of the type $\mathbf{v}[r, F_{v_1}(x)]$ where $F_{v_1}(x) = \frac{1}{lnr} \ln \frac{C(x)}{C(1)}$, the asymptotic best and worst realizations, worst-case average cost ratio, and the worst-case average cost ratio $+/-$ a standard deviation are plotted in Figure 3 (RIGHT) as a function of r. As can be seen, we can appropriately select the value of r depending on whether the goal is to minimize worst-case average cost ratio or worst-case worst-realization cost ratio. In particular, we note that by using $r = 2$, we can obtain a worst-case worst-realization cost ratio of 4, while a worst-case average cost ratio of approximately 2.8854. Therefore this particular strategy strictly outperforms the deterministic California Split search.

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

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 [7]. 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 4. In Figure 4 (LEFT), the average cost of DP(0) and RAND strategies are shown as functions of L for $\alpha = -1$, 0, and 1. In Figure 4 (RIGHT) 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.

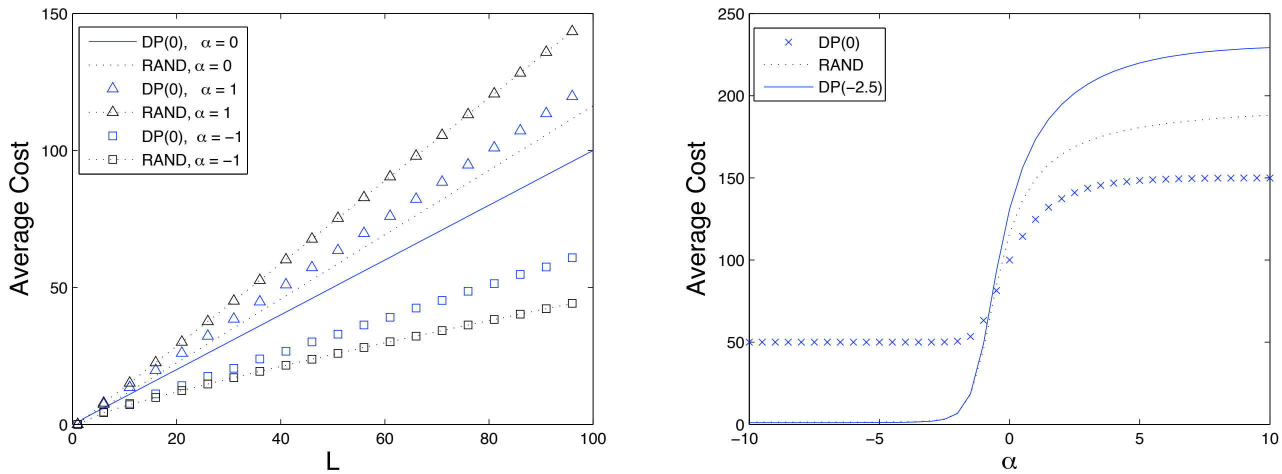

Fig. 4. (LEFT): Comparison between DP(0) and RAND for varying L and α. (RIGHT): Performance of DP(0), DP(-2.5) and RAND as functions of α when $L = 100$.

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 optimal continuous and discrete randomized strategies derived in the previous section relies on the knowledge of the functional form of the search cost $C(\cdot)$. Specifically, construction of the optimal strategy depends on the 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 a practical setting. Note that the physical meaning of search costs only exists over integer values, while continuous cost functions are introduced as a mathematical tool. If the search cost is only known for integer TTL values, then in order to obtain the optimal discrete search strategy given in Theorem 3, 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. In this case certain approximation may be used. Alternatively we could also try to develop simpler randomized strategies that are sub-optimal with respect to our performance measure but still outperform deterministic strategies and that are much easier to derive and implement than those introduced in Section V.

Motivated by this, in Section VIII we will introduce a class of such discrete randomized strategies and a number of its variations. Before we do that, we would like to 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 can be limited to the linear cost case and our presentation greatly simplified.

VII. Randomization under Linear and General Cost Functions: An Equivalence Result

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

Lemma 12: 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 $C(x) \in \mathbb{C}$. Let $\mathbf{v}$ denote the strategy that is constructed as $\mathbf{v} = C^{-1}(\mathbf{w} \cdot C(1))$, ie. $v_k = C^{-1}(w_k \cdot C(1))$ 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:

$$\sup_{x \in [1,\infty)} \frac{J_x^{\mathbf{w},l}}{x} = \sup_{y \in [1,\infty)} \frac{J_y^{\mathbf{v},g}}{C(y)} \tag{65}$$

Proof: Define the following quantities for $j \geq 1$:

$$m_j^{\mathbf{v}} = \max_{1 \leq k \leq j} \{v_k\} = \max_{1 \leq k \leq j} \{C^{-1}(w_k C(1))\} \tag{66}$$

$$m_j^{\mathbf{w}} = \max_{1 \leq k \leq j} \{w_k\} = \max_{1 \leq k \leq j} \{C(v_k)/C(1)\} , \tag{67}$$

where $m_0^{\mathbf{v}} = 0$ and $m_0^{\mathbf{w}} = 0$. Note that $C(m_j^{\mathbf{v}}) = C(1)m_j^{\mathbf{w}}$ for all j.

Fix $y \in [1, \infty)$. Note that there exists an $x \in [1, \infty)$ such that $x = C(y)/C(1)$. For finite object location y, there must exist some integer k such that the object will be located with probability 1 by the first k TTL values, ie. $Pr(m_k \geq y) = 1$. This statement is true, as explained earlier, because the strategy $\mathbf{v}$ must be able to locate the object with probability 1. Hence we have the following:

$$J_y^{\mathbf{v},g} = \sum_{j=1}^{k} E\left[C\left(v_j\right)\left|m_{j-1}^{\mathbf{v}} < y\right.\right] Pr\left(m_{j-1}^{\mathbf{v}} < y\right) = \sum_{j=1}^{k} E\left[w_j \cdot C(1)\left|m_{j-1}^{\mathbf{v}} < y\right.\right] Pr\left(m_{j-1}^{\mathbf{v}} < y\right)$$

$$= C(1)\sum_{j=1}^{k} E\left[w_j\left|C\left(m_{j-1}^{\mathbf{v}}\right) < C\left(y\right)\right.\right] Pr\left(C\left(m_{j-1}^{\mathbf{v}}\right) < C\left(y\right)\right) \tag{68}$$

$$= C(1)\sum_{j=1}^{k} E\left[w_j\left|m_{j-1}^{\mathbf{w}} < x\right.\right] Pr\left(m_{j-1}^{\mathbf{w}} < x\right) = C(1)J_x^{\mathbf{w},l} \tag{69}$$

Hence, we have the following for all $y \in [1, \infty)$:

$$\frac{J_y^{\mathbf{v},g}}{C(y)} = \frac{C(1)J_x^{\mathbf{w},l}}{C(y)} = \frac{J_x^{\mathbf{w},l}}{x}\ , \tag{70}$$

where $x = C(y)/C(1)$. Since the result holds for all $y \in [1, \infty)$ while the cost function is increasing and continuous, then (65) follows. ■

Lemma 12 implies that for any strategy $\mathbf{w}$ under linear cost there corresponds a strategy $\mathbf{v}$ that has the same performance under any cost function in $\mathbb{C}$, and vice versa. Therefore we have the following:

Theorem 5: For any two cost functions $C_1(x)$, $C_2(x)$ in $\mathbb{C}$, the best worst-case cost ratio is the same, i.e.,

$$\inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v},1}}{C_1(x)} = \inf_{\mathbf{v}\in V} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v},2}}{C_2(x)}\ , \tag{71}$$

where $J_x^{\mathbf{v},1}$ and $J_x^{\mathbf{v},2}$ are the expected search costs under $C_1(x)$ and $C_2(x)$, respectively.

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 12, the strategy $\mathbf{v} = C^{-1}\left(\mathbf{w}{\cdot}C(1)\right)$ has the same worst-case cost ratio under cost function $C(x) \in \mathbb{C}$. The k-th TTL random variable v_k therefore has probability density function f_{v_k} defined as follows for all $y \in [1, \infty)$:

$$f_{v_k}(y) = f_{w_k}\left(\frac{C(y)}{C(1)}\right) \cdot \frac{dC(y)}{dy}\frac{1}{C(1)}\ . \tag{72}$$

When v_k's are mutually independent, (72) for all k uniquely defines the strategy $\mathbf{v}$.

VIII. Uniform Randomization

In this subsection we introduce a class of *uniformly randomized strategies* and a number of its variations. Although sub-optimal, they are simple and easy to derive, and at the same time maintain the robustness of a randomized strategy. For most of this section we will limit our attention to linear cost functions, since the results can be generalized to general cost functions as discussed in the previous section. We will illustrate in Section VIII-C how our results apply to more general cost functions.

Definition 10: 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:

$$Pr(u_k = l) = \begin{cases} \frac{1}{b_{k+1}-b_k} & \text{if} \quad b_k \leq l \leq b_{k+1}-1 \\ 0 & \text{otherwise} \end{cases} \tag{73}$$

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 $\hat{g}_k$ 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.

As the constant α in the linear cost $C(k) = \alpha k$ gets cancelled out in the computation of the cost ratio, we will simply assume that the cost is $C(k) = k$ which does not affect our discussion. Then the worst-case performance measure given by (1) reduces to for any $\alpha > 0$:

$$\rho^{\mathbf{u}} = \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{x} \; . \tag{74}$$

A. Properties of Uniform Randomization

Lemma 13: Under a uniformly randomized search strategy $\mathbf{u}$ with boundaries defined by the fixed sequence $\mathbf{g}$, the worst-case cost ratio is given by:

$$\rho^{\mathbf{u}} = \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{x} = \sup_{m \in \mathbb{Z}^+} \frac{J_{b_m}^{\mathbf{u}}}{b_m} = \sup_{m \in \mathbb{Z}^+} \frac{\sum_{k=1}^{m} b_k + \frac{b_{m+1} - b_1}{2} - \frac{m}{2}}{b_m} \; . \tag{75}$$

Proof: Consider any uniformly randomized TTL sequence $\mathbf{u}$. In order to prove Lemma 13, we will first determine the possible values of x such that $\frac{J_x^{\mathbf{u}}}{x} = \rho^{\mathbf{u}}$. From (73), each expected TTL value can be calculated as:

$$E\left[u_k\right] = \frac{b_k + b_{k+1} - 1}{2} \; . \tag{76}$$

Now we can calculate $\rho^{\mathbf{u}}$. Because $\mathbf{g}$ is an increasing sequence of positive integers that increases to ∞, any positive integer x must lie between two consecutive elements of $\mathbf{g}$ such that $b_n \le x \le b_{n+1}$. Let's rewrite x as $x = b_n + \Delta$, where $0 \le \Delta \le b_{n+1} - b_n$. Then the expected cost $J_x^{\mathbf{u}}$ of using a TTL sequence $\mathbf{u}$ when the object location is x is given by:

$$J_x^{\mathbf{u}} \;=\; \sum_{k=1}^{n} E\left[u_k\right] + P(u_n < x) E\left[u_{n+1}\right] \;=\; \sum_{k=1}^{n} E\left[u_k\right] + \frac{\Delta}{b_{n+1} - b_n} E\left[u_{n+1}\right] \; .$$

We will show that the ratio $\frac{J_x^{\mathbf{u}}}{x}$ is either nonincreasing or nondecreasing for all values of x within $b_n \le x \le b_{n+1}$, and therefore the maximum value of this cost ratio within this range occurs at either $x = b_n$ or $x = b_{n+1}$.

We have for all $b_n \le x \le b_{n+1} - 1$:

$$\begin{aligned}
(x+1)J_x^{\mathbf{u}} - xJ_{x+1}^{\mathbf{u}} &= (b_n + \Delta + 1)\left(\sum_{k=1}^{n} E\left[u_k\right] + \frac{\Delta E\left[u_{n+1}\right]}{b_{n+1} - b_n}\right) - (b_n + \Delta)\left(\sum_{k=1}^{n} E\left[u_k\right] + \frac{(\Delta+1) E\left[u_{n+1}\right]}{b_{n+1} - b_n}\right) \\
&= \sum_{k=1}^{n} E\left[u_k\right] + \frac{\Delta - (b_n + \Delta)}{b_{n+1} - b_n} E\left[u_{n+1}\right] = \sum_{k=1}^{n} E\left[u_k\right] - \frac{b_n}{b_{n+1} - b_n} E\left[u_{n+1}\right] \; .
\end{aligned} \tag{77}$$

Therefore, the sign of the difference $\frac{J_x^{\mathbf{u}}}{x} - \frac{J_{x+1}^{\mathbf{u}}}{x+1} = \frac{(x+1)J_x^{\mathbf{u}} - xJ_{x+1}^{\mathbf{u}}}{x(x+1)}$ does not change for x in $b_n \le x \le b_{n+1} - 1$ because the numerator of this difference is constant (does not depend on Δ) as given by equation (77) and the denominator is always positive. Therefore, the cost ratio is either nonincreasing or nondecreasing for x in $b_n \le x \le b_{n+1}$, so the maximum cost ratio in this region occurs at either $x = b_n$ or $x = b_{n+1}$. Therefore, the maximum value of the ratio $\frac{J_x^{\mathbf{u}}}{x}$ must be obtained at $x = b_m$ for some positive integer m. In other words,

$$\rho^{\mathbf{u}} = \sup_{x \in \mathbb{Z}^+} \frac{J_x^{\mathbf{u}}}{x} = \sup_{m \in \mathbb{Z}^+} \frac{J_{b_m}^{\mathbf{u}}}{b_m} = \sup_{m \in \mathbb{Z}^+} \frac{\sum_{k=1}^{m} E\left[u_k\right]}{b_m} = \sup_{m \in \mathbb{Z}^+} \frac{\sum_{k=1}^{m} b_k + \frac{b_{m+1} - b_1}{2} - \frac{m}{2}}{b_m}$$

Therefore, we have proven Lemma 13 for any uniformly randomized strategy. ■

Lemma 13 implies that the worst-case object location for a uniformly randomized strategy must be on a boundary b_m for some m (this is the lower boundary of one of the uniform distributions), rather than an arbitrary positive integer. This greatly simplifies the process of finding the worst-case cost ratio. It also gives the expression of this cost ratio in terms of the boundary sequence.

B. Optimal Uniform Randomization

Consider the following sequence $\mathbf{b} = \{b_k\}$, $b_k = \lfloor r^{k-1} \rfloor$ for some positive real number r, $k = 1, 2, \cdots$. Define as in (73) a uniformly randomized search strategy $\mathbf{u}$ using the boundary sequence $\mathbf{b}$. Note that each $b_k = r^{k-1} - \delta_{k-1}$ for some $0 \le \delta_{k-1} < 1$.

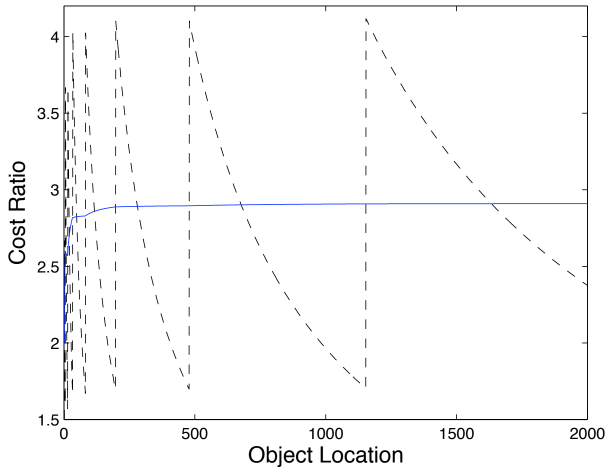

Fig. 5. Cost ratio as a function of object location for a nonrandom TTL sequence (dotted line) with $b_k = \lfloor r^{k-1} \rfloor$, $r = \sqrt{2}+1$ and the cost ratio for its uniformly randomized version (solid line). Cost is assumed to be linear.

Taking this boundary value into (75), we obtain the cost ratio for the randomized sequence:

$$\begin{aligned}\frac{J^{\mathbf{u}}_{b_m}}{b_m} &= \frac{\sum_{k=1}^{m}(r^{k-1} - \delta_{k-1}) + \frac{r^m - \delta_m - r^0 + \delta_0}{2} - \frac{m}{2}}{r^{m-1} - \delta_{m-1}} \\ &= \frac{r^{m-1}}{r^{m-1} - \delta_{m-1}}\left(\frac{r}{r-1} + \frac{r}{2} - \frac{1}{2r^{m-1}}\right) - \frac{r^{m-1}}{r^{m-1} - \delta_{m-1}}\left(\frac{\sum_{k=1}^{m}\delta_{k-1} + \frac{\delta_m + \delta_0}{2} + \frac{m}{2}}{r^{m-1}}\right)\end{aligned}$$

It can be seen from this result that for m large enough, $\frac{J^{\mathbf{u}}_{b_m}}{b_m}$ is an increasing function of m, and that we can obtain the supremum by taking the asymptotic limit:

$$\rho^{\mathbf{u}} = \sup_{m \in \mathbb{Z}^+} \frac{J^{\mathbf{u}}_{b_m}}{b_m} = \lim_{m \to \infty} \frac{J^{\mathbf{u}}_{b_m}}{b_m} = \frac{r}{r-1} + \frac{r}{2} \tag{78}$$

Differentiating (78) and noting convexity, we find that the value of r that minimizes $\rho^{\mathbf{u}}$ is $r = \sqrt{2} + 1 \approx 2.4142$, which achieves a worst-case cost ratio of $\frac{3}{2} + \sqrt{2} \approx 2.9142$. This ratio represents a 27% improvement over the worst-case cost ratio of 4 for the nonrandom California Split algorithm. The resulting uniformly randomized TTL sequence is defined by the boundary sequence $[1, 2, 5, 14, 33, \cdots]$ by taking the optimal value r into the power series.

The next theorem establishes the optimality of this uniformly randomized search strategy.

Theorem 6: Let U' denote the set of all nonrandom and uniformly randomized TTL sequences. Then:

$$\inf_{\mathbf{u} \in U'} \rho^{\mathbf{u}} = \inf_{\mathbf{u} \in U'} \sup_{x \in \mathbb{Z}^+} \frac{J^{\mathbf{u}}_x}{x} = \frac{3}{2} + \sqrt{2} \approx 2.9142\,. \tag{79}$$

That is, the uniformly randomized sequence given by the boundary sequence $b_k = \lfloor r^{k-1} \rfloor$ with $r = \sqrt{2} + 1$ is asymptotically optimal within the set U'.

The proof can be found in the appendix. Figure 5 depicts the cost ratio of using this random TTL sequence, along with the cost ratio of using its nonrandom boundary sequence b_k as TTL values. Here we observe the same qualitative difference as discussed before. Using the nonrandom TTL sequence results in oscillation of the cost ratio, while the uniformly randomized search sequence results in a smooth cost ratio curve and approaches the maximum 2.9142 asymptotically from below as the network dimension grows to infinity.

Finally, it should be noted that the above results for discrete sequences have analagous results when extending the set of admissible strategies to V. The class of strategies becomes the following:

Definition 11: Consider any infinite, increasing fixed sequence $\mathbf{b} = [b_1, b_2, ...]$ in which the elements b_k are positive real numbers (greater than or equal to 1), $b_j > b_k$ for all $j > k$, and $\lim_{k \to \infty} b_k = \infty$. 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:

$$f_{v_k}(j) = \begin{cases} \frac{1}{b_{k+1}-b_k} & \text{if} \quad b_k \leq j < b_{k+1} \\ 0 & \text{otherwise} \end{cases} \tag{80}$$

Note that Definition 11 is the continuous version of Definition 10. It can be shown that for such uniformly randomized continuous-valued TTL sequences, we have the following result which is similar to Theorem 6:

Theorem 7: Let U' denote the set of all nonrandom and uniformly randomized continuous-valued TTL sequences. Then:

$$\inf_{\mathbf{v}\in V'} \sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{x} = \frac{3}{2} + \sqrt{2} \approx 2.9142\,. \tag{81}$$

That is, the uniformly randomized sequence given by the boundary sequence $b_k = r^{k-1}$ with $r = \sqrt{2}+1$ is asymptotically optimal within the set $V^{'}$.

The proof of Theorem 7 is very similar to that of Theorem 6 and is therefore omitted.

C. Uniform Randomization for General Cost Functions

Using Lemma 12 we can obtain a discrete TTL sequence $\mathbf{u}$ which performs similarly (under any increasing cost function) as the optimal uniformly randomized sequence performed under the linear cost function (described in Section VIII-A). We first show an example when the cost is quadratic, ie $C(x) = \alpha x^2$. 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 (72) to determine the probability distribution of each TTL random variable. In particular, we have the following:

$$f_{v_k}(y) = \begin{cases} \frac{2y}{b_{k+1}^2-b_k^2} & \text{if} \quad b_k \leq y < b_{k+1} \\ 0 & \text{otherwise} \end{cases}\,, \tag{82}$$

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 why we considered the modified strategy $\hat{\mathbf{w}}$ rather than the original strategy $\mathbf{w}$. From this continuous-valued sequence, we can construct the integer-valued *discretized* version $\mathbf{u} = [u_1, u_2, ...]$ by assigning the following probability distribution to each TTL value u_k:

$$Pr(u_k = l) = \begin{cases} \int_l^{l+1} f_{v_k}(x)\,dx & \text{if} \quad b_k \leq l \leq b_{k+1}-1 \\ 0 & \text{otherwise} \end{cases}\,. \tag{83}$$

Note that this discretization essentially concentrates the probability density onto integer points, i.e. $u_k = \lfloor v_k \rfloor$ for all k. Using it with our strategy $\mathbf{v}$ in (82) gives the following:

$$Pr(u_k = l) = \begin{cases} \frac{2l+1}{{b_{k+1}}^2-{b_k}^2} & \text{if} \quad b_k \leq l \leq b_{k+1}-1 \\ 0 & \text{otherwise} \end{cases}\,, \tag{84}$$

where again $b_k = \lfloor (\sqrt{2}+1)^{\frac{k-1}{2}} \rfloor$.

The cost ratio for $\mathbf{u}$ under the quadratic cost function is depicted in Figure 6. Note that this plot is numerically very similar to Figure 5, which depicted the optimal uniformly randomized sequence under a linear cost assumption. In both curves, the randomized sequences obtain an asymptotic maximum worst-case cost of approximately 2.9142. On the other hand, if the uniform randomization of Definition 10 is applied directly to this boundary sequence under the quadratic cost function, then we obtain the dotted curve in Figure 6 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 $C(x) \in \mathbb{C}$, one can create a continuous uniformly randomized strategy $\hat{\mathbf{w}}$ with the k-th boundary value equal to $C\left(\lfloor C^{-1}\left(r^{k-1}C(1)\right)\rfloor\right)/C(1)$. The performance of this strategy under linear cost function should be similar to the optimal uniformly randomized strategy (depending on the function $C(x)$ which affects the boundary values). Then, the mapping of (72) can be used to create a strategy $\mathbf{v}$ under cost $C(x)$. Finally, apply the discretization described in (83) to this continuous strategy to obtain the discrete strategy $\mathbf{u}$, where the k-th TTL random variable will have the following distribution:

$$Pr(u_k = l) = \begin{cases} \frac{C(l+1)-C(l)}{C(b_{k+1})-C(b_k)} & \text{if} \quad b_k \leq l \leq b_{k+1}-1 \\ 0 & \text{otherwise} \end{cases}\,, \tag{85}$$

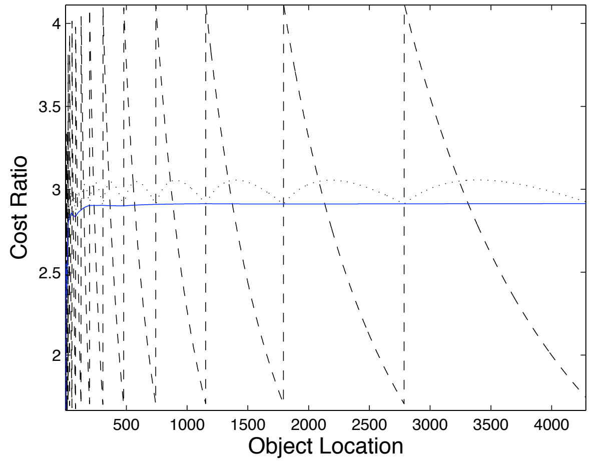

Fig. 6. Under a quadratic cost function, the cost ratio as a function of object location for a nonrandom TTL sequence (dashed line) with $b_k = \lfloor r^{\frac{k-1}{2}} \rfloor$, $r = \sqrt{2} + 1$, its uniformly randomized version (dotted line) corresponding to distribution given in (73), and its randomized version (solid line) corresponding to the distribution given in (84). Note the distribution given by (84) produces cost ratio curve that is similar to Figure 5.

where $b_k = \left\lfloor C^{-1}\left(r^{k-1}C(1)\right)\right\rfloor$. Note that while the intermediate step (mapping from $\hat{\mathbf{w}}$ to $\mathbf{v}$) requires $C(x) \in \mathbb{C}$, the final distribution in (85) does not. Therefore this method can be applied when the search cost is only defined for integer values (when $C^{-1}\left(r^{k-1}C(1)\right)$ 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 V. 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 $\hat{\mathbf{w}}$.

D. Discussion

In this subsection, we will examine the performance of uniformly randomized strategies under the measures that were described in Section VI. In the following subsection, we will describe how jointly defined TTL sequences can be used to improve performance with respect to these new criteria. We will assume in this analysis the cost is linear, but noting that due to the mapping discussed in Section VII, our results are easily extendable to general cost functions.

In general, consider any uniformly randomized strategy $\mathbf{u}$ defined by the boundary values $\mathbf{b} = [b_1, b_2, ...]$. Fix a positive integer object location x; there must exist a positive integer k such that $b_{k-1} \le x < b_k$. Then the worst-realization cost ratio is:

$$\Gamma_x^{\mathbf{u}} = \frac{1}{x}\left[\sum_{l=2}^{k-1}(b_l - 1) + x - 1 + b_{k+1} - 1\right]. \tag{86}$$

In other words, the first $k-2$ TTL values are the highest possible, the $(k-1)$th TTL value is slightly less than x, and the kth TTL value is also its highest possible. By the opposite reasoning, the best-realization cost ratio for x is when the first $k-2$ TTL values are their lowest possible, and the $(k-1)$th TTL value is equal to x. In other words,

$$\gamma_x^{\mathbf{u}} = \frac{\sum_{l=1}^{k-2} b_l + x}{x} \tag{87}$$

These numbers can be easily computed for arbitrary boundary sequences $\mathbf{b}$.

As discussed earlier, another factor to consider when analyzing any uniformly randomized strategy $\mathbf{u}$ is the cost ratio variance. Consider the same uniformly randomized strategy $\mathbf{u}$. We will use the same notation as in Section VI, as well as write $Var(Y)$ to denote the variance of any random variable Y. Fix any object location x. Then we have using independence between TTL values:

$$\Lambda_x^{\mathbf{u}} = Var\left(\sum_{k=1}^{m} u_k + I(u_m < x)u_{m+1}\right) = \sum_{k=1}^{m-1} Var\left(u_k\right) + Var(u_m + I(u_m < x)u_{m+1}) \tag{88}$$

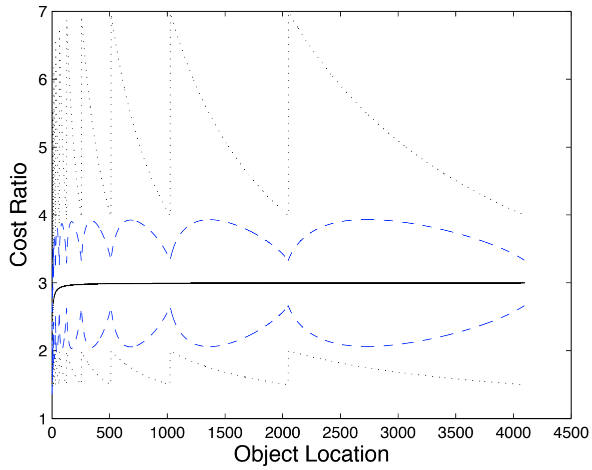

Fig. 7. Performance of uniformly randomized California Split Rule under a linear cost function. Worst and best realization cost ratio, (dotted lines), average cost ratio (solid), average cost ratio $+/-$ one standard deviation (two dashed lines).

The righthand term in (88) can be calculated as follows:

$$\begin{aligned} Var\left(u_m + I(u_m < x)u_{m+1}\right) = &\ E\left[u_m^2\right] + 2E\left[u_{m+1}\right]E[u_m I(u_m < x)] + P(u_m < x)\ \cdot E\left[u_{m+1}^2\right] \\ &- E\left[u_m\right]^2 - 2E\left[u_{m+1}\right]\ \cdot P(u_m < x)E\left[u_m\right] - P(u_m < x)^2 E\left[u_{m+1}\right]^2 \\ = &\ Var(u_m) + 2E\left[u_{m+1}\right]\frac{(x-b_m)(x-b_{m+1})}{2(b_{m+1}-b_m)}\ + P(u_m < x)\left\{Var(u_{m+1}) + P(u_m \geq x)E\left[u_{m+1}\right]^2\right\} \end{aligned} \tag{89}$$

Using this result in (88) gives us:

$$\begin{aligned} \Lambda_x^{\mathbf{u}} = \sum_{k=1}^{m} Var\left(u_k\right) + \left(b_{m+1} + b_{m+2} - 1\right)\ \cdot \frac{(x-b_m)(x-b_{m+1})}{(b_{m+1}-b_m)} \\ + \frac{x-b_m}{b_{m+1}-b_m}\left\{Var(u_{m+1}) + \frac{b_{m+1}-x}{b_{m+1}-b_m}\left(\frac{b_{m+1}+b_{m+2}-1}{2}\right)^2\right\} \end{aligned} \tag{90}$$

Finally, since u_k is uniformly distributed between b_k and $b_{k+1} - 1$, then $Var(u_k) = \frac{(b_{k+1}-b_k)^2-1}{12}$. The cost ratio variance for fixed location x is simply $\Lambda_x^{\mathbf{u}}/x^2$ when the cost is linear. Using these quantities, one can calculate the standard deviation of the cost ratio.

Figure 7 depicts the performance of the uniformly randomized California Split algorithm, as a function of object location, 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 previously discussed. The reason can be explained as follows. Since 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, such a randomization can lead to some inefficient realizations. For example, if the 5-th TTL value has realization $2^{5} - 1 = 31$, then it would be inefficient to allow the 6-th TTL value to have realization of $2^{6-1} = 32$. On the other hand, if successive TTL values are non-independent, then such inefficient realizations can be removed. Figure 8 (LEFT) depicts one example of how the probability distribution of the TTL values 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 $0 \leq \delta \leq 2^{k-1} - 1$, 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 8 (RIGHT) 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 8 (RIGHT) and 7, it can be seen that the cost ratio for the tree construction has less deviation from its mean value.

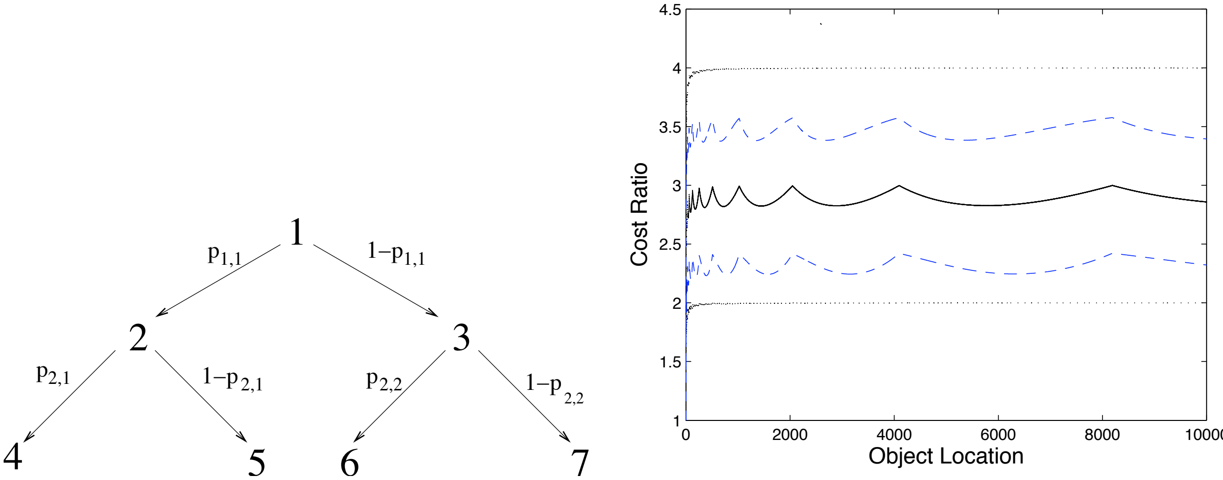

Fig. 8. (LEFT): Example of how a binary tree can be used to construct a TTL sequence. In particular, the figure indicates that our first TTL value is 1. With probability $p_{1,1}$, the second TTL value will be 2 and with probability $1-p_{1,1}$, the second TTL value will be 3. If the second TTL value is 2, then with probability $p_{2,1}$ the third TTL value will be 4, and with probability $1-p_{2,1}$ it will be equal to 5. Likewise, if the second TTL value is 3, then with probability $p_{2,2}$ the third TTL value will be 6, and with probability $1-p_{2,2}$ it will be equal to 7. This process can be extended to construct an infinite-length TTL sequence. (RIGHT): Performance of randomization proposed by (LEFT) figure if $p_{i,j} = \frac{1}{2}$ 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 8 (LEFT) in order 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.

IX. Conclusion and Future Work

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 provided a randomization construction for decreasing the worst-case cost ratio of a given fixed strategy. We also derived an asymptotically optimal strategy whose search cost is always within a factor of e of the cost of an omniscient observer. We provided a mapping between TTL sequences under different cost functions, and then derived the optimal strategy within the class of uniformly randomized strategies for linear search cost. These results are directly applicable in designing practical algorithms.

References

[1] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in the sensor networks," *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networks (MobiCOM'99)*, August 1999, Seattle, Washington.
[2] D. Braginsky and D.Estrin, "Rumor routing algorithm for sensor networks," *Proc. International Conference on Distributed Computing Systems (ICDCS-22)*, 2002.
[3] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, pp. 153–181, 1999.
[4] 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.
[5] 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.
[6] S. Shakkottai, "Asymptotics of query strategies over a sensor network," *Proceedings of IEEE Infocom*, March 2004, Hong Kong.
[7] 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.
[8] 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.
[9] 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.
[10] A. Borodin and R. El-Yaniv, *Online Comptutation and Competetive Analysis*, Cambridge University Press, 1998, Cambridge, UK.
[11] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, 1996, New York, USA.

APPENDIX

A. Proof of Lemma 3

We prove this Lemma by contradiction. Suppose for some x^* satisfying equation (12) that the claim is not true, which means that either $x^* = u_n$ or $x^* = u_n + a < u_{n+1}$ for some $a \geq 2$ and for some $0 \leq n \leq N-1$. We will prove the contradiction for both cases.

Case 1: Suppose $x^* = u_n$ for $1 \leq n \leq N-1$. Then the corresponding search cost $J^{\mathbf{u}}_{x^*} = \sum_{l=1}^{n} C(u_l)$. This can be rearranged as:

$$\frac{J^{\mathbf{g}}_{u_n}}{C(u_n)} = \frac{\sum_{l=1}^{n} C(u_l)}{C(u_n)} = \frac{\frac{C(u_n+1)}{C(u_n)}\left(\sum_{l=1}^{n-1} C(u_l) + C(u_n)\right)}{C(u_n+1)}, \tag{91}$$

where the value of the summation on the righthand side of (91) is equal to zero if $n = 1$. However, using $m = u_n + 1$ in the constraint of equation (11), along with the fact that $\sum_{l=1}^{n-1} C(u_l) \leq \sum_{i=1}^{n-1} C(i)$, and then rearranging gives us:

$$\frac{C(u_n+1)}{C(u_n)} \sum_{l=1}^{n-1} C(u_l) < C(u_n) + \sum_{l=1}^{n-1} C(u_l) = \sum_{l=1}^{n} C(u_l) \tag{92}$$

In addition, because $u_n + 1 \leq u_{n+1}$, then it follows that $J^{\mathbf{u}}_{u_n+1} = \sum_{l=1}^{n+1} C(u_l) \geq \sum_{l=1}^{n} C(u_l) + C(u_n+1)$. Combining equations (91) and (92) and using this inequality gives us:

$$\frac{J^{\mathbf{u}}_{u_n}}{C(u_n)} < \frac{\sum_{l=1}^{n+1} C(u_l)}{C(u_n+1)} \leq \frac{J^{\mathbf{u}}_{u_n+1}}{C(u_n+1)} \tag{93}$$

However, this contradicts the assumption that $x^* = u_n$ satisfies $\frac{J^{\mathbf{u}}_{x^*}}{C(x^*)} = \max_{1 \leq x \leq L} \frac{J^{\mathbf{u}}_x}{C(x)}$. Therefore it cannot be true that $x^* = u_n$ for some $1 \leq n \leq N-1$.

Case 2: Now we can consider the second case of $x^* = u_n + a < u_{n+1}$ for $1 \leq n \leq N-1$ and some $a \geq 2$. Then $J^{\mathbf{u}}_{x^*} = \sum_{l=1}^{n+1} C(u_l)$. However, we also have $J^{\mathbf{u}}_{u_n+1} = \sum_{l=1}^{n+1} C(u_l)$. This gives us:

$$\frac{J^{\mathbf{u}}_{x^*}}{C(x^*)} = \frac{\sum_{l=1}^{n+1} C(u_l)}{C(u_n+a)} < \frac{\sum_{l=1}^{n+1} C(u_l)}{C(u_n+1)} = \frac{J^{\mathbf{u}}_{u_n+1}}{C(u_n+1)} \tag{94}$$

Again, this contradicts the assumption that x^* satisfies $\frac{J^{\mathbf{u}}_{x^*}}{C(x^*)} = \max_{1 \leq x \leq L} \frac{J^{\mathbf{u}}_x}{C(x)}$. Therefore, it cannot be true that $x^* = u_n + a < u_{n+1}$ for $0 \leq n \leq N-1$ and $a \geq 2$ ∎

B. Proof of Theorem 1

To begin, we will use the notation that m is in the set R if $u_m + 1 \in S$, for any $0 \leq m \leq N-1$. Hence R has at most N members, and each member is less than N. Next, note that for any $1 \leq x \leq L$, there must exist a corresponding positive integer m such that $u_{m-1} < x \leq u_m$ (because the TTL sequence is strictly increasing and $u_N = L$). Then for $x \notin S$, the corresponding cost of the randomized sequence is given by $J^{\hat{\mathbf{u}}}_x = J^{\mathbf{u}}_x + pM_{u_m+1}$, where we define $M_{L+1} = M_L$ for notational reasons. This statement is true because for any such x:

$$J^{\hat{\mathbf{u}}}_x = p\left(\sum_{k \in R,\ 0 \leq k \leq m} C(u_k+1) + \sum_{k \notin R,\ 1 \leq k \leq m} C(u_k)\right) + (1-p)\sum_{k=1}^{m} C(u_k) = \sum_{k=1}^{m} C(u_k) + pM_{u_m+1} = J^{\mathbf{u}}_x + pM_{u_m+1}$$

Now we will prove Theorem 1 for two separate cases.

Case 1: $u_N \notin S$.

This case corresponds to the sequence generated by either (C.4) or (C.5). We will prove that $\frac{J^{\hat{\mathbf{u}}}_x}{x} < \rho^{\mathbf{g}}$ for all x. First, let's consider all $x \notin S$. As stated earlier, $J^{\hat{\mathbf{u}}}_x = J^{\mathbf{u}}_x + pM_{u_m+1}$. We then have from inequality (14) defining our chosen p:

$$\frac{J^{\hat{\mathbf{u}}}_x}{C(x)} = \frac{J^{\mathbf{u}}_x + pM_{u_m+1}}{C(x)} \leq \frac{J^{\mathbf{u}}_x}{C(x)} + \frac{pM_L}{C(x)} < \frac{J^{\mathbf{u}}_x}{C(x)} - \frac{J^{\mathbf{u}}_x}{C(x)} + \rho^{\mathbf{u}} = \rho^{\mathbf{u}}$$

Therefore, $\frac{J^{\hat{\mathbf{u}}}_x}{C(x)} < \rho^{\mathbf{u}}$ for all $x \notin S$.

If $1 \in S$, then it must be true from Lemma 3 that $u_1 > 1$. In addition, $J_1^{\hat{\mathbf{u}}} = pC(1) + (1-p)C(u_1) < C(u_1) = J_1^{\mathbf{u}}$, where the strict inequality holds because the cost function is strictly increasing and $u_1 > 1$. Therefore the cost ratio has decreased at location 1.

Next consider the case $x = u_m + 1 \in S$ for some positive integer m. This means that the expected search cost is given by:

$$\begin{aligned} J_x^{\hat{\mathbf{u}}} &= p\left(\sum_{k\in R,\ 0\le k\le m} C(u_k+1) + \sum_{k\notin R,\ 1\le k\le m} C(u_k)\right) + (1-p)\sum_{k=1}^{m+1} C(u_k) \\ &= \sum_{k=1}^{m+1} C(u_k) + pM_{u_m} - pC(u_{m+1}) < \sum_{k=1}^{m+1} C(u_k) = J_x^{\mathbf{u}} \end{aligned} \tag{95}$$

where the last inequality in (95) follows from the fact that $M_{u_m} \le C(u_m) < C(u_{m+1})$, which follows from the definition of M_j. Equation (95) implies that $\frac{J_x^{\hat{\mathbf{u}}}}{C(x)} < \frac{J_x^{\mathbf{g}}}{C(x)} = \rho^{\mathbf{u}}$ because $x \in S$ and achieves the maximum cost ratio for $\mathbf{u}$.

Combining the above, we have that $\frac{J_x^{\hat{\mathbf{g}}}}{C(x)} < \rho^{\mathbf{g}}$ for all integers $1 \le x \le L$ when $u_N \notin S$.

Case 2: $u_N \in S$.

This case corresponds to (C.6) and (C.7). We first consider $1 \le x < u_{N-1} = u_N - 1$. For these values of x, we have that $\frac{J_x^{\hat{\mathbf{u}}}}{C(x)} < \rho^{\mathbf{u}}$ by following the same steps used in the first part (case 1) of this proof. As discussed earlier, if $u_N \in S$ then this means that $u_{N-1} = u_N - 1$. In addition, from Lemma 3 we know that $u_{N-1} \notin S$. Therefore, when $x = u_{N-1}$, we have:

$$J_x^{\hat{\mathbf{u}}} = (1-p)\sum_{k=1}^{N-1} C(u_k) + p\left(\sum_{k\in R,\ 0\le k\le N-2} C(u_k+1) + \sum_{k\notin R,\ 1\le k\le N} C(u_k)\right) = J_x^{\mathbf{u}} + pM_L$$

which gives:

$$\frac{J_x^{\mathbf{u}} + pM_L}{C(x)} < \frac{J_x^{\mathbf{u}}}{C(x)} - \frac{J_x^{\mathbf{u}}}{C(x)} + \rho^{\mathbf{u}} = \rho^{\mathbf{u}}\ , \tag{96}$$

where the last inequality follows from inequality (14) defining our chosen p. Since $u_N = u_{N-1} + 1$ is the only value of x such that $x > u_{N-1}$, it only remains to prove that $\frac{J_{u_N}^{\hat{\mathbf{u}}}}{C(u_N)} < \rho^{\mathbf{u}}$. When $x = u_N$, we have the following expected search cost:

$$J_L^{\hat{\mathbf{u}}} = \sum_{k=1}^{N-2} C(u_k) + pM_{L-2} + (1-p)C(u_{N-1}) + C(u_N) = \sum_{k=1}^{N} C(u_k) + p(M_{L-2} - C(u_N - 1)) < \sum_{k=1}^{N} C(u_k) = J_L^{\mathbf{u}}\ ,$$

where the last inequality follows from the fact that $C(u_{N-1}) > M_{L-2}$ using the definition of M_j and fact that $u_{N-1} = L-1$.

Combining these two cases, we have that $\frac{J_x^{\hat{\mathbf{u}}}}{C(x)} < \rho^{\mathbf{u}}$ for all $1 \le x \le L$ and have proven this theorem. ∎

C. Proof of Theorem 6

It has been shown in [8] that the maximum cost ratio for any nonrandom TTL strategy is bounded below by 4, and therefore to calculate the infimum given in (79), we need to only consider uniformly randomized strategies. We will prove Theorem 6 by showing that $\frac{3}{2} + \sqrt{2}$ is both a lower bound and an upper bound on $\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}}$.

We begin by showing that $\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}} \ge \frac{3}{2} + \sqrt{2}$. We will proceed using proof by contradiction via a similar method to the one presented in [8] to establish the lower bound on the maximum cost ratio for any nonrandom TTL strategy. Assume that the maximum cost ratio for a uniformly randomized sequence $\mathbf{u}$, defined by the boundary values $\mathbf{b} = [b_1, b_2, ...]$, is some constant $\vartheta < \frac{3}{2} + \sqrt{2}$. We have already shown that the worst-case ratio for $\mathbf{u}$ takes the form given in (78). Therefore, by this equation and the assumption that the maximum ratio is ϑ, then the following must be true for all $m \in \mathbb{Z}^+$:

$$\sum_{k=1}^{m} b_k + \frac{b_{m+1} - b_1}{2} - \frac{m}{2} \le \vartheta b_m \implies \sum_{k=1}^{m} b_k + \frac{b_{m+1}}{2} \le \vartheta b_m + B_m$$

where $B_m = \frac{b_1}{2} + \frac{m}{2}$. Now define $\tilde{y}_n = \sum_{k=1}^{n} b_k$, so the above equation becomes:

$$\tilde{y}_m + \frac{1}{2}(\tilde{y}_{m+1} - \tilde{y}_m) \le \vartheta(\tilde{y}_m - \tilde{y}_{m-1}) + B_m \implies \tilde{y}_{m+1} + (1-2\vartheta)\tilde{y}_m + 2\vartheta\tilde{y}_{m-1} \le 2B_m \tag{97}$$

Now, because $\mathbf{b}$ is an increasing sequence of positive integers, $\tilde{y}_m$ is increasing faster than B_m. This fact means that for some $N \geq 0$, we have: $\tilde{y}_{N+1} > B_{N+1} + \frac{\vartheta}{2} - \frac{1}{4}$. Let $y_k = \tilde{y}_{N+k} - B_{N+k} - \frac{\vartheta}{2} + \frac{1}{4}$, so that the y_k are increasing and positive on $\mathbb{Z}^+$. Our above equation then becomes under this new variable with $m = N + k$:

$$y_{k+1} + B_{N+k+1} + \frac{\vartheta}{2} - \frac{1}{4} + (1-2\vartheta)\left(y_k + B_{N+k} + \frac{\vartheta}{2} - \frac{1}{4}\right) + 2\vartheta\left(y_{k-1} + B_{N+k-1} + \frac{\vartheta}{2} - \frac{1}{4}\right) \leq 2B_{N+k}$$

Rearranging, we obtain:

$$y_{k+1} + (1-2\vartheta)y_k + 2\vartheta y_{k-1} \leq (2\vartheta+1)B_{N+k} - (1-2\vartheta)B_{N+k+1} - \vartheta + \frac{1}{2}$$

Using the definition of $B_m = \frac{b_1}{2} + \frac{m}{2}$, we obtain:

$$y_{k+1} + (1-2\vartheta)y_k + 2\vartheta y_{k-1} \leq \frac{N+k}{2}(2\vartheta+1) - \frac{N+k+1}{2} - \frac{N+k-1}{2}2\vartheta - \vartheta + \frac{1}{2}$$

Cancelling out terms, we obtain:

$$y_{k+1} + (1-2\vartheta)y_k + 2\vartheta y_{k-1} \leq 0 \tag{98}$$

Now, we will prove that (98) cannot hold for all k if $\vartheta < \frac{3}{2} + \sqrt{2}$. Form a sequence $...\xi_{-1}, \xi_0 = 0, \xi_1 = 1, \xi_2, ...$ which satisfies the equation $\xi_{l-1} + (1-2\vartheta)\xi_l + 2\vartheta\xi_{l+1} = 0$. Note that this sequence is uniquely defined by its values $\xi_0 = 0$ and $\xi_1 = 1$. Then the corresponding characteristic equation for this sequence is:

$$1 + (1-2\vartheta)\lambda + 2\vartheta\lambda^2 = 0 \tag{99}$$

The nature of the roots of this characteristic equation can be determined by calculating $(1-2\vartheta)^2 - 4(2\vartheta) = 4\vartheta^2 - 12\vartheta + 1$. Notice that for $\vartheta = \frac{3}{2} + \sqrt{2}$, we have $4\vartheta^2 - 12\vartheta + 1 = 0$ and that for $1 \leq \vartheta < \frac{3}{2} + \sqrt{2}$, it is always true that $4\vartheta^2 - 12\vartheta + 1 < 0$. In the latter case, the characteristic equation has complex conjugate roots which means that the solution to $\xi_{l-1} + (1-2\vartheta)\xi_l + 2\vartheta\xi_{l+1} = 0$ has a sinusoidal form. Therefore, there exists some $M \geq 1$ such that $\xi_i > 0$ for $0 < i < M+1$ and that $\xi_{M+1} \leq 0$. Also we know that $\xi_{-1} < 0$ from the recursive equation defining our sequence. So from equation (98), we have:

$$\sum_{i=1}^{M} \left(y_{i+2} + (1-2\vartheta)y_{i+1} + 2\vartheta y_i\right)\xi_i \leq 0 \tag{100}$$

This equation can be arranged into the following:

$$\sum_{i=1}^{M+1} y_i\left(\xi_{i-2} + (1-2\vartheta)\xi_{i-1} + 2\vartheta\xi_i\right) + \left[-2\vartheta y_{m+1}\xi_{m+1} - (1-2\vartheta)y_1\right] + \left[y_{m+2}\xi_m - y_2\xi_0 - y_1\xi_{-1}\right] \leq 0 \tag{101}$$

However, the first term above is zero by the recursive equation for our ξ_i, and the second and third terms (to the left of the inequality) are both positive due to the fact that $\xi_{-1} < 0, \xi_0 = 0, \xi_1 = 1, \xi_m > 0, \xi_{m+1} < 0$ and $y_i > 0$ for all i. Therefore, we have arrived at a contradiction and it cannot be possible that $\vartheta < \frac{3}{2} + \sqrt{2}$. Hence, $\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}} \geq \frac{3}{2} + \sqrt{2}$.

However, we have already shown that for the uniformly randomized sequence $\mathbf{u}$ defined by the boundary values $b_k = \lfloor r^{k-1} \rfloor$ where $r = \sqrt{2} + 1$, the worst-case cost ratio $\rho^{\mathbf{u}}$ is $\frac{3}{2} + \sqrt{2}$. It thus follows that $\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}} \leq \frac{3}{2} + \sqrt{2}$.

Combining these two results, we see that $\inf_{\mathbf{u}\in U'} \rho^{\mathbf{u}} = \frac{3}{2} + \sqrt{2}$. ∎

D. *Proof of Lemma 6*

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 (20), the ratio between cost of successive TTL values can be expressed in terms of $C(v_1)$ as follows for any integer $j \geq 1$:

$$\frac{C(v_{j+1})}{C(v_j)} = \left(\alpha^{-\sum_{k=0}^{j-1}\alpha^k}\right)\left(\frac{C(v_1)}{C(1)}\right)^{\alpha^j} \tag{102}$$

Consider any positive finite j. If (26) holds then we have by using (102):

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

which holds for all integers j. Hence, (26) 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 (102) and using $C(v_{j+1}) > C(v_j)$:

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

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 (26) is also a necessary condition for an increasing sequence. ∎

E. Proof of Lemma 8

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 $k \geq 1$. Note that the expected value of S_k can be calculated as follows:

$$E\left[S_k\right] = E\left[\sum_{j=1}^{k} C(v_j)\right] = E\left[\sum_{j=1}^{k} r^{j-1}C(v_1)\right] = \sum_{j=0}^{k-1} r^j E\left[C(v_1)\right] = E\left[C(v_1)\right]\sum_{j=0}^{k-1} r^j = E\left[C(v_1)\right]\frac{r^k-1}{r-1} \tag{103}$$

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

$$\begin{aligned} E\left[C(v_1)|v_1 \leq l\right] &= \int_0^{\infty} Pr\left(C(v_1) > y|v_1 \leq l\right)\, dy = C(1) + \int_{C(1)}^{\infty} \frac{Pr\left(C^{-1}(y) < v_1 \leq l\right)}{Pr(v_1 \leq l)}\, dy \\ &= C(1) + \frac{1}{F_{v_1}(l)}\left[\int_{C(1)}^{C(l)} \left[\bar{F}_{v_1}(C^{-1}(y)) - \bar{F}_{v_1}(l)\right]\, dy\right] \\ &= \frac{1}{F_{v_1}(l)}\left[C(1) + \int_{C(1)}^{C(l)} \bar{F}_{v_1}(C^{-1}(y))\, dy - C(l)\cdot\bar{F}_{v_1}(l)\right] \end{aligned} \tag{104}$$

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 in other words $C^{-1}\left(r^{n-1}C(1)\right) \leq x < C^{-1}\left(r^n C(1)\right)$. Then the expected search cost $J_x^{\mathbf{v}}$ can be calculated as follows by using (103):

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

Using (104), we obtain the following:

$$\begin{aligned} J_x^{\mathbf{v}} &= \frac{r^n-1}{r-1}E\left[C(v_1)\right] + r^n \cdot \left[C(1) + \int_{C(1)}^{\frac{C(x)}{r^{n-1}}} \bar{F}_{v_1}(C^{-1}(y))\, dy - \frac{C(x)}{r^{n-1}}\bar{F}_{v_1}\left(C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right)\right] \\ &= \frac{r^n}{r-1}\left[E\left[C(v_1)\right] + (r-1)\left\{C(1) + \int_{C(1)}^{\frac{C(x)}{r^{n-1}}} \bar{F}_{v_1}\left(C^{-1}(y)\right)\, dy\right\}\right] - rC(x)\bar{F}_{v_1}\left(C^{-1}\left(\frac{C(x)}{r^{n-1}}\right)\right) - \frac{E\left[C(v_1)\right]}{r-1} \end{aligned} \tag{105}$$

Letting $z = \frac{C(x)}{r^{n-1}C(1)}$, we obtain the following expression for the cost ratio by plugging into (105):

$$\begin{aligned} \frac{J_x^{\mathbf{v}}}{C(x)} &= \frac{r}{(r-1)}\frac{E\left[C(v_1)\right] + (r-1)\left\{C(1) + \int_{C(1)}^{z\cdot C(1)} \bar{F}_{v_1}(C^{-1}(y))\, dy\right\}}{zC(1)} - r\frac{\bar{F}_{v_1}(C^{-1}(z))}{C(1)} - \frac{E\left[C(v_1)\right]}{(r-1)zr^{n-1}} \\ &= \frac{r}{r-1}\frac{h(r) + (r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} - \frac{h(r)}{(r-1)zr^{n-1}C(1)} \end{aligned} \tag{106}$$

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

$$\Phi_n\left(z\right) = \frac{r}{r-1}\frac{h(r)+(r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)} - \frac{h(r)}{(r-1)zr^{n-1}C(1)}\ , \tag{107}$$

so that from equation (106), $\Phi_n(z)$ is simply the cost ratio at object location $x = C^{-1}\left(zr^nC(1)\right)$. Note that 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 first two terms in the expression for the cost ratio in (106) are the same for x and y, and the third term increases with increasing x. In addition, when x ranges from $C^{-1}(r^{n-1}C(1))$ to $C^{-1}(r^nC(1))$, 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 approaches ∞ of $\Phi_n(z)$ is simply $\Phi(z)$, where $\Phi(z)$ is the function defined earlier in (30). Hence, the following is true, where $x_n = C^{-1}(r^nC(1))$ for notation:

$$\begin{aligned}
\sup_{x\in[1,\infty)} \frac{J_x^{\mathbf{v}}}{C(x)} &= \sup_{n\in\mathbb{Z}^+}\left\{\sup_{x_{n-1}\le x<x_n}\frac{J_x^{\mathbf{v}}}{C(x)}\right\} = \sup_{n\in\mathbb{Z}^+}\left\{\sup_{1\le z<r}\Phi_n\left(z\right)\right\} \\
&= \sup_{1\le z<r}\left\{\sup_{n\in\mathbb{Z}^+}\Phi_n\left(z\right)\right\} = \sup_{1\le z<r}\left\{\lim_{n\to\infty}\Phi_n\left(z\right)\right\} \\
&= \sup_{1\le z<r}\left\{\Phi(z)\right\} = \sup_{1\le z<r}\left\{\frac{r}{r-1}\frac{h(r)+(r-1)h(z)}{zC(1)} - r\frac{h'(z)}{C(1)}\right\}\ ,
\end{aligned} \tag{108}$$

which completes the proof of the lemma. ∎

F. Proof of Lemma 9

Because $\frac{C(x+1)}{C(x)} \ge q$ for all x, we have that $\frac{C(X)_\alpha+1}{C(X_\alpha)} \ge q,\ \ w.p.1$. Hence, $\frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} \ge q$ for all α. Therefore to complete the proof, we need to show that $\lim_{\alpha\to1^+}\frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} \le q$. This is equivalent to showing 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 $1 < \alpha < \bar{\alpha}$.

Fix $\epsilon > 0$. Since $\lim_{x\to\infty}\frac{C(x+1)}{C(x)} = q$, there exists a x^* such that $\frac{C(x+1)}{C(x)} < q + \frac{\epsilon}{2}$ for all $x > x^*$. Let $1(\cdot)$ denote the indicator function; so $1(A) = 1$ if A is true, otherwise it equals 0. Thus we have:

$$E[C(X_\alpha+1)1(X_\alpha > x^*)] <\ \left(q+\frac{\epsilon}{2}\right)E[C(X_\alpha)1(X_\alpha > x^*)] \le\ \left(q+\frac{\epsilon}{2}\right)E[C(X_\alpha)] \tag{109}$$

At the same time, we have:

$$\lim_{\alpha\to1^+}\frac{E[C(X_\alpha+1)1(X_\alpha \le x^*)]}{E[C(X_\alpha)]} \le\ \lim_{\alpha\to1^+}\frac{C(x^*+1)}{E[C(X_\alpha)]} = 0\ ,$$

because $C(x^*+1) < \infty$ and $E[C(X_\alpha)] = \frac{\alpha}{\alpha-1}$, which approaches ∞ as α goes to 1. Hence, there exists an $\bar{\alpha}$ such that for all$1 < \alpha < \bar{\alpha}$:

$$\frac{E[C(X_\alpha+1)1(X_\alpha \le x^*)]}{E[C(X_\alpha)]} < \frac{\epsilon}{2} \tag{110}$$

Therefore, combining (109) and (110) gives for all $1 < \alpha < \bar{\alpha}$

$$\frac{E[C(X_\alpha+1)]}{E[C(X_\alpha)]} = \frac{E[C(X_\alpha+1)1(X_\alpha > x^*)]}{E[C(X_\alpha)]} + \frac{E[C(X_\alpha+1)1(X_\alpha \le x^*)]}{E[C(X_\alpha)]} <\ q + \frac{\epsilon}{2} + \frac{\epsilon}{2} = q+\epsilon\ ,$$

which completes the proof. ∎

G. Variance of Optimal Continuous Strategy

Now we examine the variance of the strategy $\mathbf{v}[r,\ F_{v_1}(x)]$ where $F_{v_1}(x) = (ln\, C(x)/C(1))/ln\, r$. To begin, let $f_{v_1}\left(x\,|v_1 < y\right)$ denote the pdf of v_1 given that v_1 is less than y. Note the following for $1 \le y < C^{-1}\left(rC(1)\right)$

$$f_{v_1}\left(x\,|v_1 < y\right) = \frac{dF_{v_1}\left(x\,|v_1 < y\right)}{dx} = \begin{cases} \frac{1}{C(x)\ ln\ [C(y)/C(1)]}\frac{dC(x)}{dx} & \text{if}\ \ 1 \le x < y \\ 0 & \text{otherwise} \end{cases} \tag{111}$$

Then we have:

$$E\left[C\left(v_1\right)^2\middle|\, v_1 < y\right] = \int_1^y \frac{C(x)^2}{C(x) ln\, C(y)/C(1)} \frac{dC(x)}{dx}\, dx = \frac{C(y)^2 - C(1)^2}{2 \cdot ln\, C(y)/C(1)} \tag{112}$$

It can easily shown that:

$$E\left[C\left(v_1\right)^2\right] = \frac{C(1)^2}{ln\, r}\left[\frac{r^2-1}{2}\right] \tag{113}$$

Finally, using the above two results gives us the following:

$$E\left[C\left(v_1\right)^2\middle|\, v_1 \geq y\right] = \frac{E\left[C(v_1)^2\right] - E\left[C\left(v_1\right)^2\middle|\, v_1 < y\right] Pr\left(v_1 < y\right)}{Pr(v_1 \geq y)} = \frac{C(1)^2 r^2 - C(y)^2}{2\left(ln\, r - ln\, C(y)/C(1)\right)} \tag{114}$$

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 $J_x^{\mathbf{v}} = E\left[j_x^{\mathbf{v}}\right]$. The second moment of this search cost can be calculated as follows by using (112), (113), and (114):

$$\begin{aligned}
E\left[\left(j_x^{\mathbf{v}}\right)^2\right] &= E\left[\left(j_x^{\mathbf{v}}\right)^2\middle|\, v_k < x\right] Pr\left(v_k < x\right) + E\left[\left(j_x^{\mathbf{v}}\right)^2\middle|\, v_k \geq x\right] Pr\left(v_k \geq x\right) \\
&= E\left[\left(\sum_{l=1}^{k+1} r^{l-1} C(v_1)\right)^2\middle|\, v_k < x\right] Pr\left(v_k < x\right) + E\left[\left(\sum_{l=1}^{k} r^{l-1} C(v_1)\right)^2\middle|\, v_k \geq x\right] Pr\left(v_k \geq x\right) \\
&= \left(\frac{r^{k+1}-1}{r-1}\right)^2 E\left[C(v_1)^2\middle|\, v_1 < C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right] Pr\left(v_1 < C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right) \\
&\quad + \left(\frac{r^k-1}{r-1}\right)^2 E\left[C(v_1)^2\middle|\, v_1 \geq C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right] Pr\left(v_1 \geq C^{-1}\left(\frac{C(x)}{r^{k-1}}\right)\right) \\
&= \frac{(r^{k+1}-1)^2\left[\left(\frac{C(x)}{r^{k-1}}\right)^2 - C(1)^2\right] + (r^k-1)^2\left[C(1)^2 r^2 - \left(\frac{C(x)}{r^{k-1}}\right)^2\right]}{2(ln\, r)(r-1)^2}
\end{aligned} \tag{115}$$

In addition, it can be easily shown that:

$$J_x^{\mathbf{v}} = \frac{rC(x) - C(1)}{ln\, r} \tag{116}$$

Note that the variance of the cost ratio at location x is simply the difference between (115) and the square of (116), divided by $C(x)^2$. Hence we have:

$$\begin{aligned}
\frac{\Lambda_x^{\mathbf{v}}}{C(x)^2} &= \frac{(r^{k+1}-1)^2\left[\left(\frac{C(x)}{r^{k-1}}\right)^2 - C(1)^2\right] + (r^k-1)^2\left[C(1)^2 r^2 - \left(\frac{C(x)}{r^{k-1}}\right)^2\right]}{2(ln\, r)(r-1)^2 C(x)^2} - \left(\frac{rC(x) - C(1)}{ln\, r}\right)^2 \frac{1}{C(x)^2} \\
&= \frac{(r^{k+1}-1)^2\left[r^{-2k+2} - \frac{C(1)^2}{C(x)^2}\right] + (r^k-1)^2\left[\frac{C(1)^2}{C(x)^2} r^2 - r^{-2k+2}\right]}{2(ln\, r)(r-1)^2} - \left(\frac{r - \frac{C(1)}{C(x)}}{ln\, r}\right)^2
\end{aligned} \tag{117}$$

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

$$\lim_{x\to\infty} \frac{\Lambda_x^{\mathbf{v}}}{C(x)^2} = \lim_{k\to\infty} \frac{(r^{k+1}-1)^2\left[r^{-2k+2}\right] - (r^k-1)^2 r^{-2k+2}}{2(ln\, r)(r-1)^2} - \left(\frac{r}{ln\, r}\right)^2 = \frac{r^4 - r^2}{2(ln\, r)(r-1)^2} - \frac{r^2}{(ln\, r)^2} \tag{118}$$