On search sets of expanding ring search in wireless networks

By: Jing Deng , Sergei Zuyev

J. Deng and S. Zuyev, "On Search Sets of Expanding Ring Search in Wireless Networks," *Elsevier Ad Hoc Networks*, vol. 6, no. 7, pp. 1168-1181, September 2008.

Made available courtesy of Elsevier: http://dx.doi.org/10.1016/j.adhoc.2007.11.003

Abstract

We focus on the problem of finding the best search set for expanding ring search (ERS) in wireless networks. ERS is widely used to locate randomly selected destinations or information in wireless networks such as wireless sensor networks In ERS, controlled flooding is employed to search for the destinations in a region limited by a time-to-live (TTL) before the searched region is expanded. The performance of such ERS schemes depends largely on the search set, the set of TTL values that are used sequentially to search for one destination. Using a cost function of searched area size, we identify, through analysis and numerical calculations, the optimum search set for the scenarios where the source is at the center of a circular region and the destination is randomly chosen within the entire network. When the location of the source node and the destination node are both randomly distributed, we provide an almost-optimal search set. This search set guarantees the search cost to be at most 1% higher than the minimum search cost, when the network radius is relatively large.

Keywords: Flooding; Expanding ring search; Wireless networks; Search cost; Optimum

1. Introduction

In wireless communication networks such as wireless ad hoc networks [7] and wireless sensor networks [1], network nodes may need to inquire destinations that are unknown to themselves thus far. Since such an inquiry may take place before a routing path is found, routing information is usually unavailable in the context. One way to send the inquiry packet is to use the flooding technique, in which the packet is broadcasted and each neighboring node forwards the packet once. The process continues until every node in the network has forwarded the packet once. Such a flooding scheme is usually termed as pure flooding. Pure flooding is rather expensive since it involves all nodes in the network. In fact, a “broadcast storm" problem [12] may appear when inappropriate rebroadcasts are performed and when packet collisions occur frequently, requiring more rebroadcasts. These collisions degrade the overall network performance and should be avoided. Advanced techniques have been developed to reduce the number of redundant rebroadcasts while maintaining the reachability of the flooding process [12].

Since the destination may reside in an area that is relatively close to the source node, an expanding ring search (ERS) technique is employed in a number of networking protocols, such as routing [9,13] and information query [10]. In the ERS scheme, the query packet is broadcasted with a time-to-live (TTL) value. When the packet is received by other nodes, the TTL value on the packet is decremented. Then the packet will be rebroadcasted only if the TTL value is positive. The exact implementations of the ERS scheme vary in different protocols. For instance, in some ERS techniques, the initial TTL value is set to 1 for the first (ring) search. If such a search fails to find the destination, a new search is initiated with an incremented TTL value 2. The process continues until the initial TTL value reaches a threshold, L. Then a network-wide flooding is initiated [8]. The ERS scheme in dynamic source routing (DSR) [9] uses the California split rule [2], where the TTL value is doubled every time when the previous TTL value fails. In ad hoc on-demand distance vector (AODV) routing [13], an ERS scheme is implemented to start with a TTL value of TTL_START and to increase the TTL by TTL_INCREMENT after each failure.

Recently, several research papers have been published with the focus of analyzing search strategies such as the ERS schemes under various network settings [5,3,8]. It has been shown that the ERS schemes that increment the TTL value by one after each failure are generally ineffective. Therefore, it is interesting to identify the best search strategy, in terms of which set of TTLs should be used, for ERS. In this work, we analyze the cost of different search sets in the ERS technique and investigate the optimum search sets. We develop a general analytical framework to measure the search cost of different search sets in the ERS schemes. Based on this framework, we study wireless networks from which a destination is randomly chosen. When the source is at the origin of the circular network region, we identify the optimum search sets. When the source is randomly chosen from the network region, we provide an almost-optimal search set, which guarantees the search cost to be at most 1% higher than the minimum search cost when the network radius is relatively large.

Similarly to other researchers [5,3,4], we neglect the effect of underlying medium access control (MAC) schemes and broadcast collisions in our analysis. Our focus in this work is on first-order effect of different search sets. We do not consider second-order effects such as those caused by different MAC protocols. We further argue that such collisions and, when necessary, rebroadcasts, will only change the search cost proportionally. Using this approach enables us to focus on the search cost caused by different search sets.

The difference of our work with related work [5,3,4,8] is that our work focus on identifying a search set that performs close to the optimum search set. We develop a general analytical framework to find such search sets. The framework can serve as a tool for further study in this field. We also provide guidelines for search set selections under various network setups.

The rest of this paper is organized as follows. In Section 2, we summarize the related work. In Section 3, we present the network model of our analysis and our analytical framework to investigate the optimum search sets in ERS. We study the scenarios where the source is at the origin of the circular network region in Section 4. The scenarios in which the source is randomly chosen from the circular region are studied in Section 5. Section 6 presents numerical results on integer search sets. We conclude our work and state future work directions in Section 7.

2. Related work

Cheng and Heinzelman investigated geography- based and hop-based flooding control methods [5]. It was proved that two-tier and three-tier hop-based flooding control methods can reduce the cost of broadcast. Both of the cost and the latency were studied. A general formula to determine good parameters for two-tier and three-tier schemes was provided and investigated. Different to their work, we provide a general analytical framework and use it to investigate networks with different source distributions.

Chang and Liu revisited the TTL-based controlled flooding search extensively [3]. When the probability distribution of the location of the searched object is known *a priori*, a dynamic programming formulation of the optimal search sets can be used. They also presented the necessary and sufficient conditions on the location distribution under which pure flooding and incremental ERS became optimum, respectively. One other major result is the optimization of worst-case search cost when the probability distribution of the location of the object is unknown. The optimization problem was later extended to include delay constraints [4]. The primary goal is to derive search strategies that minimize worst-case search cost subjected to a worst-case delay constraint. Different from their work, we try to obtain optimum search set when the node distribution probability function is known *a priori*.

Hassan and Jha studied the optimum L for one class of ERS schemes with a search set of $\{1; 2,. .. , L\}$. The authors argued that there existed an optimum value of L, which minimizes the expected cost of broadcasts [8].

However, it is generally non-optimal to use a consecutive sequence of integers as the search set, as shown by other researchers [5,3].

Baryshnikov et al. studied the California split rule in a network that reduces to a path [2]. In the California split rule, the set of TTLs is $\mathbf{u} = \{x_1;x_2;...$ ~ where $x_i=2^{i-1}$ for all $i \geq 1$. In [11], alternative query algorithms of the Gnutella network were explored through simulations. A new query algorithm based on multiple random walks was proposed and evaluated.

Krishnanmachari and Ahn optimized the number of data replicas for event information in wireless sensor networks [10]. The closed-form approximations for expected energy costs of search and data replication were used to derive replication strategies that minimize cost. It was found that the number of replication should be proportional to the square root of query rates.

In our prior work [6], we generalized pure flooding and ERS into ring search (RS) schemes and developed a theoretical framework to show that a special three-step RS scheme is optimum among all three-step RS schemes.

In this paper, we develop a general analytical framework to study different search sets with different source locations. This framework is used to investigate the optimum search set when the source is at the origin of the circular network region and when the source is randomly chosen in the network region. Instead of looking for optimum search sets, we identify three-ring almost-optimal search sets with search cost guarantees (of at most 1% higher than the optimum search sets). These integer search sets can be easily applied to large wireless networks.

3. Analytical model

3.1. Assumptions and notations

We assume that a source node S is searching for a destination node, D, in a network region $X \subseteq \mathbb{R}^2$. The events of these two nodes' distribution in the network are mutually independent. We further assume that the wireless transmission media is broadcast in nature, i.e., when a node sends out a message, all physical neighbors will overhear it unless packet collisions occur. Throughout the paper, we use the following notations:

- R: the network radius;
- ξ: the distance between S and D, S — D ;
- r_i: search set for ERS, $1 \ll i \ll N$, where N is the size of the search set[1];
- λ: node density of the network; and
- $|b|$: the area of a ball b.

3.2. Analytical model

Consider a homogeneous ensemble of nodes in the plane $\mathbb{R}^2$ with large density A. Mathematically, it can be modeled as the homogeneous Poisson process with spatial density A, in other words, a homogeneous intensity measure $\Lambda(dx) = \lambda dx$, in a domain X $\mathbb{R}^2$ which is assumed to contain the origin 0 for definitiveness. Although we do not consider it here, a non-homogeneous intensity measure can be used to model non-evenly spread nodes or to take into account a known prior distribution of the destination location. Changing if necessary the metric unit, we may assume that all the nodes have the maximal communicating distance of 1 unit. In order to find the destination, the source node employs the ERS mechanism, when first only the nodes at the distance not exceeding r_1 are requested, next, at the distance not exceeding r_2, etc., until the destination is found (cf. Fig. 1). The search cost of ERS is proportional to the number of nodes involved in sending the search packets until the destination is found as it is reception handling and re-transmissions of the search packets by these nodes that distinguish the idling system from the one with an active search. Before we establish an analytic formula for the cost, we should remark the following.

Overall cost of the system functioning has many components, not all of which are relevant to the optimization of the search cost. First of all, this concerns the constant power drainage due to idling of the nodes. Next, whatever search mechanism is employed, it takes the same amount of energy to communicate back the location of the destination, once it is found, to the search originating node, so this cost may be excluded from the cost function to be optimized. Finally, the cost of re-broadcasts due to collisions of packets can be taken into account by adjusting the parameters of the transmission and reception cost, simply by multiplying them by 1 plus the probability of re-broadcast. Also, without loss of generality, the cost of handling of the packet can be included to the cost of reception.

We will be interested in the case when the density λ is high, and call the *asymptotic search* cost the almost sure limit of the normalized search cost when $\lambda \to \infty$, provided it exists. Such a situation is typical in the sensor networks in contrast to other wireless networks which may and usually are rather dense. Ever lowering cost of mass production of sensors allows for increased redundancy provided by a dense distribution of the sensors in failure critical situations like hazard monitoring or military applications. Another scenario which falls into our framework is when the network is sparse, but the communicating distance is rather large relative to the typical inter-node distances. In essence, we require that the number of nodes in direct communication of a node to be large.

When λ is high, the nodes at the distance of at most n hops from S fill the ball b_n of radius n (assuming $b_n \subset X$ for the moment). The number of nodes inside a ball of radius n is Poisson($\lambda\pi n^2$)-distributed, so, divided by λ, it approaches with probability 1 the area πn^2 with the error of order O($\lambda^{-1/2}$) that provides the basis for our choice of the cost function below.

Let P_r; P_t denote the cost of reception and transmission of the search packet, respectively. To discover the destination when $\xi \le 1$, it only takes one inquiry transmission from the source, consecutive reception of the nodes within the ball b_1. Thus the consumed energy, divided by $\lambda \gg 1$, gives the asymptotic cost approaching a constant $P_r\,|b_1| = \pi P_r$. We will ignore this constant value contribution to the cost function in the sequel. When $1 < \xi \le r_1 + 1$, then (up to the specified ball approximation error), the search cost involves requesting all the nodes in the overlap region of X and the ball b_{r1} to broadcast the inquiry once. These broadcasts are received by the nodes in one-neighborhood of the broadcasting nodes. Therefore, the incurred cost writes

$$P_t|X \cap b_{r_1}| + P_r|X \cap b_{r_1+1}|. \qquad (1)$$

If $r_1 + 1 < \xi \le r_2 + 1$, so that the first search cannot locate the destination, but the second one will, the cost will be

$$\begin{aligned} &P_t|X \cap b_{r_1}| + P_r|X \cap b_{r_1+1}| + P_t|X \cap b_{r_2}| + P_r|X \cap b_{r_2+1}| \\ &\quad = (P_r + P_t)\Big[(|X \cap b_{r_1}| + |X \cap b_{r_2}|) + \frac{P_r}{P_r + P_t} \\ &\qquad \times (|X \cap (b_{r_1+1} \setminus b_{r_1})| + |X \cap (b_{r_2+1} \setminus b_{r_2})|)\Big]. \end{aligned}$$

In general, the normalized by $1/(P_r+P_t)$ cost is equal to

$$\begin{aligned} &\mathscr{C}_0(r_1, r_2, \ldots, r_N) = \sum_{k=1}^{\cdot\cdot} [|X \cap b_{r_k}| + \alpha|X \cap \\ &\quad (b_{r_k+1} \setminus b_{r_k})|]\mathbb{1}_{\xi > r_{k-1}+1}, \qquad (2) \end{aligned}$$

where $\alpha = P_r=(P_r + P_t)$ and $r_0 = 0$ by definition (c.f. [3, Eq. (1)]). N here is the maximal number of tries, which, in principal, may be infinity. There are a few conditions to meet in order to avoid degenerate cases. The first evident condition is that

$$0 < r_1 < \cdots < r_N. \qquad (3)$$

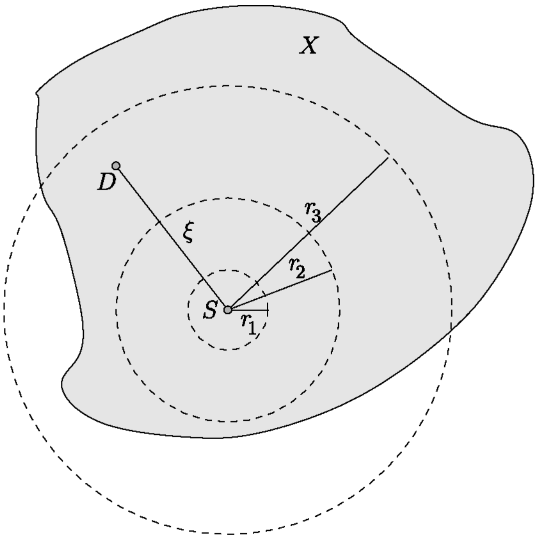

Fig. 1. Illustration of the expanding ring search (ERS) scheme. Node S looks for node D through three searches. The first two controlled flooding with radii of r_1 and r_2 fail to find D, but the third search with radius of r_3 finds it. Note that the incurred search cost is the overlapped region of the circles and the network region X.

Further conditions depend on the support of the distribution of ξ. If it is unbounded, we necessarily have $N = \infty$ and $r_n \longrightarrow \infty$. Otherwise with positive probability some distant destinations that occur with a positive probability will never be detected. If sup n is a compact set then we require that

$$\mathbf{P}\{\xi > r_{N-1} + 1\} > 0 \quad \text{and} \quad \mathbf{P}\{\xi > r_N + 1\} = 0 \tag{4}$$

for $1 \leq N < \infty$. Condition (4) states that the very last search should be flooding to all the nodes where the destination maybe found. Without loss of generality, r_N maybe put to ess sup ξ - 1, where
ess sup ξ = inf{z : P{ξ > z} = 0}.

Taking expectation with respect to the radius distribution n of the destination, we arrive at the following formula for the average cost of the search scheme using the sequence $\{r_1; r_2; \dots ; r_N\}$:

$$\begin{aligned}\mathbf{E}\mathscr{C}_0 &= \sum_{k=1}^{N} [|X \cap b_{r_k}| + \alpha |X \cap (b_{r_k+1} \setminus b_{r_k})|] \mathbf{P}\{\xi > r_{k-1} + 1\} \\ &= \sum_{k=1}^{N} [|X \cap b_{r_k}| + \alpha |X \cap (b_{r_k+1} \setminus b_{r_k})|] f(r_{k-1}) \\ &= \overline{\mathscr{C}} + \overline{\mathscr{C}}_1,\end{aligned}$$

where

$$f(z) = \mathbf{P}\{\xi > z + 1\}, \tag{5}$$

$$\overline{\mathscr{C}} = \sum_{k=1}^{N} |X \cap b_{r_k}| f(r_{k-1}), \tag{6}$$

$$\overline{\mathscr{C}}_1 = \alpha \sum_{k=1}^{N} |X \cap (b_{r_k+1} \setminus b_{r_k})| f(r_{k-1}). \tag{7}$$

The goal is to find the minimum of this cost over $0 < r_1 < r_2 < \cdots < r_N =$ ess sup ξ - 1 when ess sup $\xi < \infty$ or over an infinite growing sequence $\{r_i\}$; $i = 1; 2; \dots$ otherwise.

Before we proceed, we make the final assumption. The parameter $\alpha = P_r/(P_r + P_t)$ above does not exceed 1/2 because $P_r \leq P_t$. In fact, the cost of reception has been demonstrated by other researchers to be similar to that of the idling nodes, so α is actually a small number. In addition, the summands in (6) are proportional to the area of balls b_{rk}, so to r_k^2, while the terms in (7) behave as r_k. Since all $r_k > 1$, we argue that disregarding the smaller order term $\overline{\mathscr{C}}_1$ will not change much the optimal solution for $\mathbf{E}\mathscr{C}_0$. We will make remarks on the actual error bounds of this approximation later when we consider particular examples. From now on, we will be dealing with optimization of the goal function $\overline{\mathscr{C}}$ given by formula (6).

There may be two approaches to optimize the cost: static and dynamic. In the former case one looks for the sequence r_k (or the counting measure it) minimizing $\overline{\mathscr{C}}$. In the later case, each next r_k is chosen recursively on the basis of already tried r_1; ... ; r_{k-1}. We focus on static search set in this work.

4. Source at the origin of a circle

In this section, we assume that the network region is a circle and S is at the origin of the circle. Based on this assumption, X is in fact a ball with radius of R, b_R. The distribution function of D becomes

$$f(z) = 1 - \frac{(z+1)^2}{R^2} \quad \text{when } 0 \leqslant z \leqslant R-1 \tag{8}$$

and 0, otherwise. Here expression (6) turns into

$$\frac{1}{\pi}\overline{\mathscr{C}} = \frac{1}{R^2}\sum_{k=1}^{N} r_k^2[R^2 - (r_{k-1}+1)^2]. \tag{9}$$

Therefore, we need to look for an optimum N and $t_1, t_2, \ldots, t_{N-1}$ that minimize function

$$F_N(t_1, \ldots t_{N-1}) = \sum_{k=1}^{N} r_k^2[R^2 - (r_{k-1}+1)^2], \tag{10}$$

where $r_k = \sum_{i=1}^{k} t_i$ and $1 \leqslant k \leqslant N-1$, over the simplex

$$t_1 + \cdots + t_{N-1} \leqslant R - 1, \tag{11}$$

$$t_1, \ldots, t_{N-1} \geqslant 0. \tag{12}$$

There are two special cases. When N = 1, r0 = 0 (flooding),

$F_1\ (x_1) = F_1\ (R - 1) = (R - 1)^2(R^2 — 1)$

$= R^4 - 2R^3 + 2R - 1 = F_1^*$.

When $N = 2$, $r0 = 0$, and rN = $r2 = R — 1$, we have

$F_2(x_1) = x_1^2(R^2 — 1) + (R - 1)2(R^2 - (x^1 + 1)^2)$,

which attains the minimal value

$$F_2^* = \frac{1}{2}(2R^4 - 5R^3 + 3R^2 + R - l) \tag{13}$$

for $x_1 = (R - 1)/2$. It is seen that $F_1^* - F_2^* = (R - 1)^{3/}2 > 0$ for $R > 1$. Therefore, when $R > 1$, two-step search set always outperforms the pure flooding scheme. Since

$$\lim_{R\to\infty} \frac{F_1^* - F_2^*}{F_2^*} = 0,$$

the performance gain of using two-step search set as compared to flooding diminishes with the increase of R. When $N = 3$, (9) becomes

$$\frac{1}{\pi}\overline{\mathscr{C}} = \frac{1}{R^2}[r_1^2(R^2-1) + r_2^2(R^2 - (r_1+1)^2) + (R-1)^2(R^2 - (r_2+1)^2)], \tag{14}$$

where we have used $r3 = R — 1$.

Taking partial derivative of (14) over r1 and r2 and letting it equal to 0, respectively, we have

$$\begin{cases} r_1 = \frac{r_2^2}{R^2 - r_2^2 - 1}, \\ r_2 = \frac{(R-1)^2}{R^2 - (r_1+1)^2 - (R-1)^2}. \end{cases} \tag{15}$$

The optimum search sets with N = 3 should satisfy (15). Closed-form solutions to (15) are rather complex, however its asymptotic solutions can be found as follows:

It is very likely that r*1 « R, then r*2 satisfies

$$\lim_{R\to\infty} \frac{r_2^*}{R} = \frac{1}{2} \tag{16}$$

and r*1 should satisfy

$$\lim_{R\to\infty} r_1^* = \frac{1}{3}. \tag{17}$$

Combining (16) and (17) with (15), asymptotically r_1^* and r_2^* should be

$$r_1^* = \frac{1}{3} \quad r_2^* = \frac{(R-1)^2}{2R - \frac{25}{9}}, \tag{18}$$

which has been confirmed by our numerical results with the use of fmincon in Matlab.[2]

When $N > 3$, a closed-form expression for the optimum search sets is very difficult to obtain. We rely on numerical methods to find the optimum search set, with the use of fmincon in Matlab. The optimum search set and the search cost are shown in Figs. 2 and 3.

In Fig. 2, we show the optimum search sets r_1^* and r_2^*, with $N = 3$, as their ratio over the asymptotic values suggested in (18). Note that when $R < 13$, the optimum $N = 3$ search sets degenerate to two rings.

The search costs of different search sets are compared in Fig. 3. We *define relative* cost of a search set as its search cost normalized by that of the optimum search set, i.e., the minimum search cost:

$$\rho = \frac{\overline{\mathscr{C}}(r_1, r_2, \ldots, r_N)}{\overline{\mathscr{C}}^*} \geqslant 1, \tag{19}$$

where $\overline{\mathscr{C}}^*$ represents the minimum search cost. Based on Fig. 3 we can see that the three-step search set as defined in (18) performs very close to the optimum search set. In fact, the normalized search cost is mostly less than 1.01. When R is relatively large, flooding should be used instead since it only introduces less than 1% extra search cost.

5. Source randomly chosen from X

In Section 4, we considered the scenarios where the source node S is always at the origin of the circular region. In this section, S is assumed to be randomly distributed in the network region $X = b_R$. The problem of finding a common ERS strategy for all the nodes which would minimize the overall expected cost is on order.

Since both nodes S and D are uniformly distributed in $X \subseteq \mathbb{R}^2$ independently, the average asymptotic ring search cost may be calculated from (6):

$$\overline{\mathscr{C}}(r_1,r_2,\dots,r_N)=\mathbf{E}\sum_{k=1}^{N}|X\cap b(S,r_k)|\mathbb{1}_{\|S-D\|>r_{k-1}+1}. \quad (20)$$

As above, rN should correspond to flooding which means r_N = diam$X$— 1 = sup{||jx—y||; x;y ∈ X} - 1 and, as before, we neglect the contribution of the smaller order term (7) to the optimum solution of the cost.

The above expression simplifies to

$$\begin{aligned}\overline{\mathscr{C}}&(r_1,r_2,\dots,r_N)\\ &=\frac{1}{|X|^2}\sum_{k=1}^{N}\int\int_{X^2}|X\cap b(x_0,r_k)|\mathbb{1}_{\|x_0-x_1\|>r_{k-1}+1}\,\mathrm{d}x_0\,\mathrm{d}x_1\\ &=\frac{1}{|X|^2}\sum_{k=1}^{N}\int_X|X\cap b(x_0,r_k)|\cdot|X\setminus b(x_0,r_{k-1}+1)|\mathrm{d}x_0.\end{aligned}$$

In particular, if X is the disk b_R of radius R at the origin, then

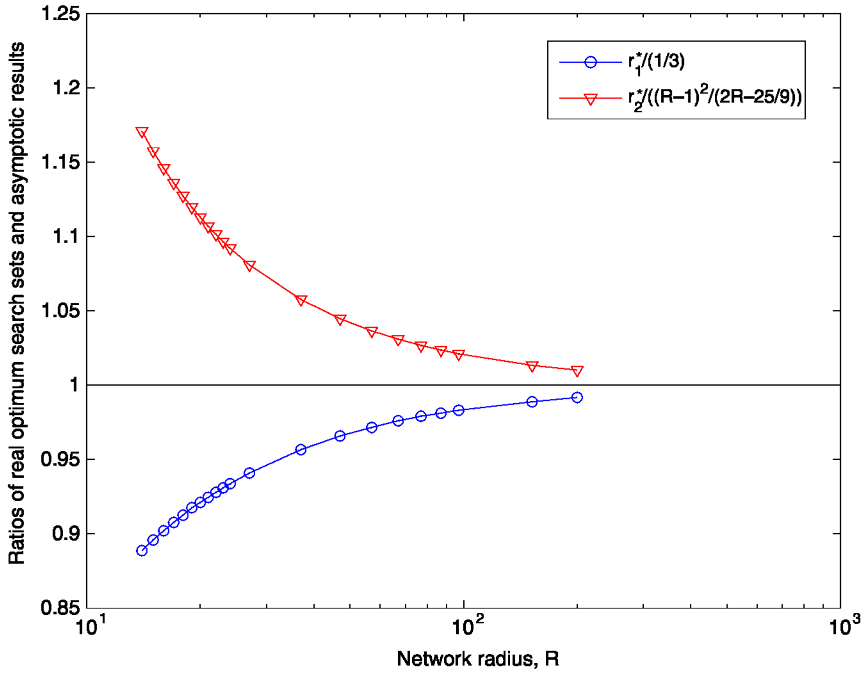

Fig. 2. Optimum search sets when $N=3$. We show r_1^* and r_2^* in terms of the ratio over the asymptotic values as suggested by (18). $r_3^*=R-1$.

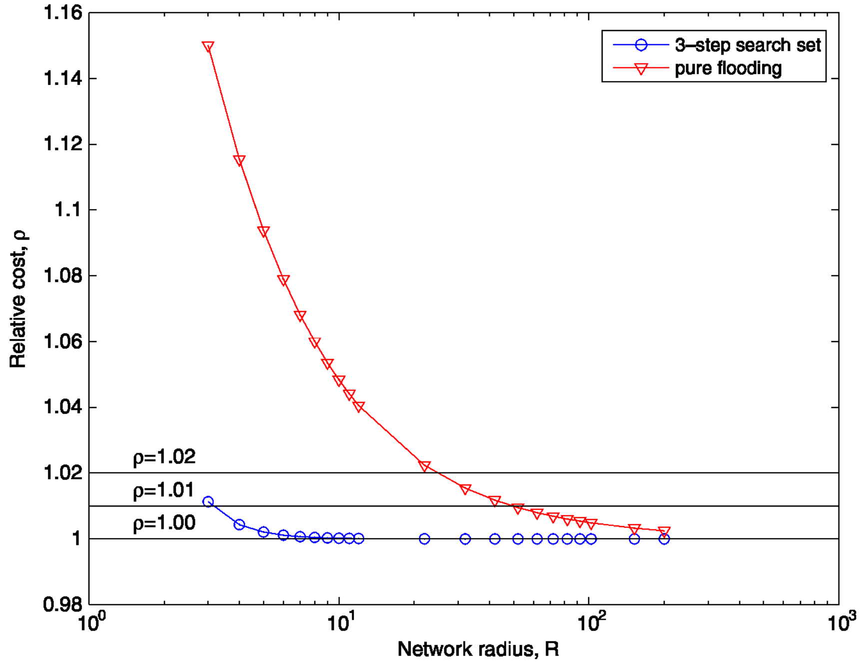

Fig. 3. Numerical results comparing the search costs of the three-step search sets as in (18) and $r_3 = R-1$, and pure flooding. The search costs have been normalized over the search cost of the optimum search sets, i.e., (19).

$$\begin{aligned}\overline{\mathscr{C}}(r_1, r_2, \ldots, r_N) &= \frac{2}{\pi R^4}\sum_{k=1}^{N}\int_0^R |b_R \cap b((\rho,0), r_k)| \\ &\quad \cdot |b_R - |b_R \cap b((\rho,0), r_{k-1}+1)|\rho\,\mathrm{d}\rho \\ &= 2\pi\sum_{k=1}^{N} I(r_{k-1}+1, r_k; R), \end{aligned} \tag{21}$$

where

$$I(s,t;R) = \int_0^R \left(1 - \frac{A(R,s,\rho)}{\pi R^2}\right)\frac{A(R,t,\rho)}{\pi R^2}\rho\,\mathrm{d}\rho. \tag{22}$$

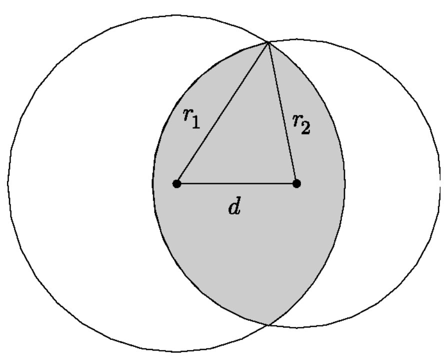

Fig. 4. Illustration of $A(r_1, r_2, d)$.

The function $A(r_1, r_2, d)$ above is the area of the intersection on two disks with radii r_1 and r_2 with their centers separated by the distance d (cf. Fig. 4). Its explicit formula is rather cumbersome, but not difficult to compute:

$$\ell = \frac{r_2^2 + d^2 - r_1^2}{2d},$$
$$A_1 = \frac{\pi r_1^2}{2} - (d-\ell)\sqrt{r_1^2 - (d-\ell)^2} - r_1^2 \arcsin\frac{d-\ell}{r_1},$$
$$A_2 = \frac{\pi r_2^2}{2} - \ell\sqrt{r_2^2 - \ell^2} - r_2^2 \arcsin\frac{\ell}{r_2},$$
$$A(r_1, r_2, d) = \begin{cases} 0 & \text{if } d \geqslant r_1 + r_2 \\ \pi r_1^2 & \text{if } r_2 \geqslant d + r_1 \\ \pi r_2^2 & \text{if } r_1 \geqslant d + r_2 \\ A_1 + A_2 & \text{otherwise} \end{cases}.$$

Typical form of functions $\overline{\mathscr{C}}$ are shown in Fig. 5.

Numeric analysis with the help of statistical computing package R[3] shows that the optimal N can be as large as 7, especially for relatively small R. For instance, for R = 1 the minimal cost of 0.3023865 is obtained for the following sequence of $r1$, ... , $r6$: 0.01092865, 0.1092967, 0.3564751, 0.6603114, 0.8808269, 0.9783151 ($r7 = r_N$ = 1). However, the minimal cost when using N = 1, ... , 6 rings are as follows: 1.842555 (flooding), 0.3756568, 0.3178889, 0.3050929, 0.3026054, 0.30245560, so we see that rather good approximation (the cost less than 1% higher than the overall minimum) to the minimum cost is already obtained when using just three rings before flooding of radii 0.1733899, 0.467521 and 0.7702083 (N = 4). Although we have not included the number of rings or latency in our cost function, practical observations suggest that one can easily tolerate an error of a few percent to decide in favor of an “almost optimal" policy with a fewer number of rings. Taking this approach with tolerance of 1% or less, we come to the following recommendation. For the ranges of R smaller than 3.4, three rings and then flooding ERS strategy should be used. When R is between 3.4 and 11, 2 rings and flooding ERS is optimal. One ring of size approximately 0.83R and flooding is optimal for mid-range $11 \leq R \leq 90$. For larger R, flooding should be used from the start, as it makes only less than 1% difference with the minimum search cost.

The best search sets, among all search sets with N = 12, 3, 4}, relative to R are shown in Fig. 6. We present the best search sets among all N = 12, 3, 41 search sets as thick lines in different segments of R. Thin lines are the best search set values of each individual N = 2, 3, and 4, respectively. For instance, for R = 5, the optimal strategy is to use N = 3 and $r_1 = 0.546R = 2.73$, $r_2 = 1.358R = 6.79$, and $r_3 = 9$ (points on the short-dashed thick lines multiplied by R = 5). Should, however, one wants to use N = 2, the search set should be $r_1 = 0.793R$ (the fourth line from bottom) which provides the minimum cost over all N = 2 search sets. Similarly, for the N = 4 case, the minimum is provided by the rings of sizes r_1 = 0.099R, $r_2 = 0.701R$, $r_3 = 1.461R$, and $r_4 = 2R - 1$, respectively (thin lines continued from the thick lines optimal in the range 1–3.4 of R).

Finally, our calculations show that adding the term $\overline{\mathscr{C}}_1$ defined in (7) even with maximal possible value of a = 1/2 altered the optimal values for the radii only slightly. For instance, when R = 10, the difference does not exceed 0.3. When α = 0. 1, the difference was negligible. This reinforces our claim that the term $\overline{\mathscr{C}}_1$ can be disregarded in the optimization goal function.

6. Integer search sets

In the protocols currently used, only integer TTL values are allowed in ERS schemes. We investigate such optimum integer search sets through numerical calculations. We used integer argument fmincon in Matlab to look for the optimum search sets.

6.1. Source at the origin of a circle

Table 1 presents the optimum integer search sets for some values of R. Our first observation is that all of these optimum search sets have only two elements, i.e., N 1/4 2. Similar results were reported in [6].[4] Based on the numbers in Table 1, we conclude that the optimum search sets are $\{r_1^*, r_2^*\} = \{\lfloor (R-1)/2 \rfloor, R-1\}$

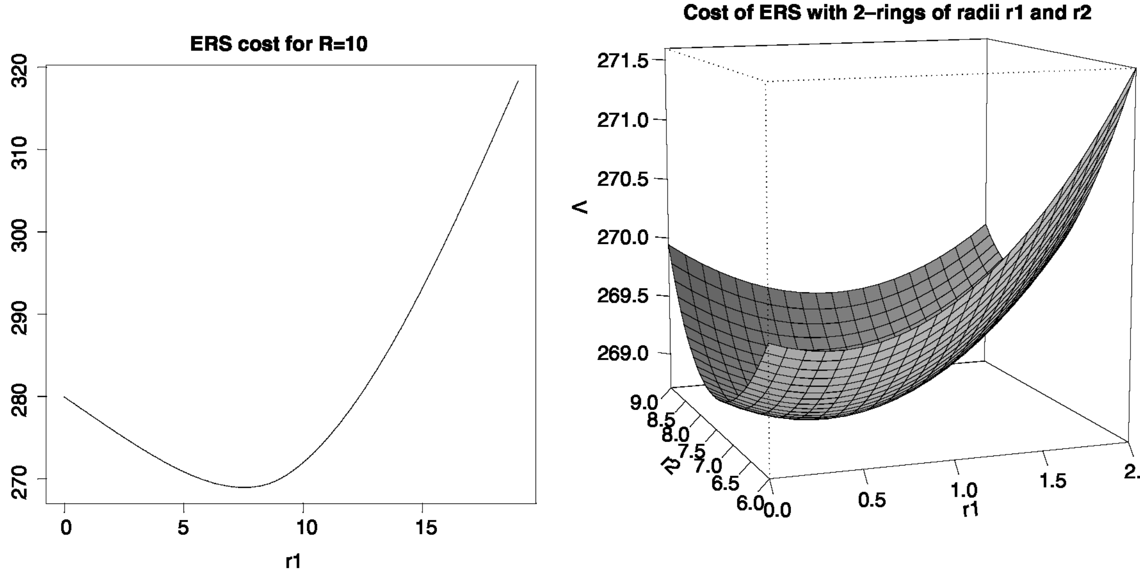

Fig. 5. Typical cost functions for one (left graph) and two rings used prior to flooding.

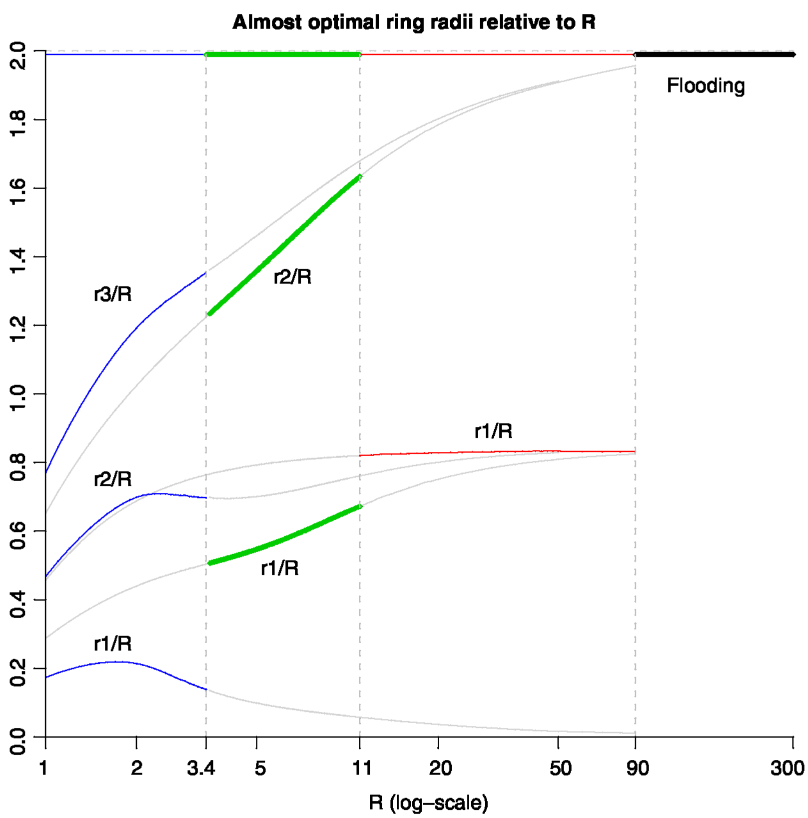

Fig. 6. The best search sets among all search sets with $N = \{2,3,4\}$. The thick lines of different styles in each region of R represent the best search sets among all $N = \{2,3,4\}$ search sets: solid thick lines for $R \leqslant 3.4$, thick short-dashed for $3.4 < R \leqslant 11$, and long-dashed for $11 < R < 90$. The thin lines are extension of the thick lines into other regions of R values.

In Fig. 7, we present the normalized search cost, p, the ratio between the search cost of flooding and the minimum search cost in each R. We can see from Fig. 7 that, as R increases, the cost saving with the use of the optimum search set is diminishing. In fact, when R = 50, the cost saving of using the optimum search set as compared to using flooding is roughly 1%. With the consideration of search delay and additional operational costs, flooding should be chosen instead as R increases further.

Table 1
Optimum integer search sets for different R

R	3	4	5	6	10	15	20	25	30	35	40	45	50
r_1^*	1	2	2	2	4	7	9	12	14	17	19	22	24
r_2^*	2	3	4	5	9	14	19	24	29	34	39	44	49

There are exactly two elements in these optimum sets. The source is assumed to be at the origin of a circular network region, $X = b_R$.

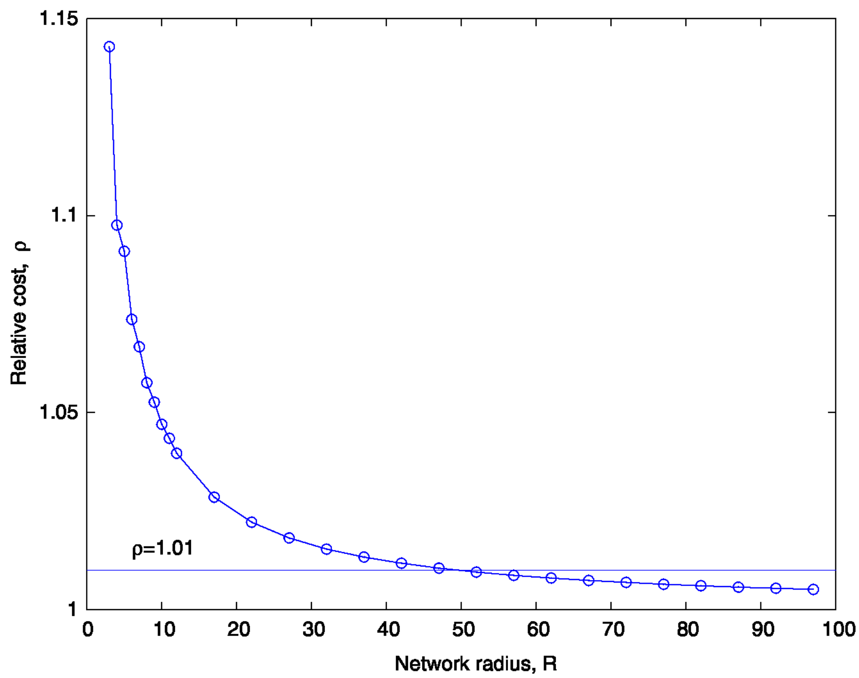

Fig. 7. Relative cost of pure flooding. The sources are assumed to be fixed at the origin of a circle.

All results shown above are based on the assumption of dense networks, which should satisfy (8). It would be interesting to see how our scheme reacts to networks that do not satisfy (8). In Fig. 8, we compare the relative cost of our suggested search set and the pure flooding scheme under sparser networks through Matlab simulations. The source is assumed to be at the origin of the circles with radius of 4 or 6. We can observe from Fig. 8 that our scheme operates with a search cost that is at most 10% higher than that of the optimum search set. The pure flooding scheme has much higher relative cost.

In order to investigate the performance of our suggested search set in sparse and small-diameter networks, we simulated our suggested search set and pure flooding scheme in networks with radius of 1, 2, and 3. The results are presented in Fig. 9. N changes from 4 to 30 with node density being as low as 0.14 node per unit area. When radius is 3, the relative cost decreases as N increases. Therefore, the benefit of using our suggested search set improves as node density increases. An interesting fluctuation of relative cost in small N range can be seen for networks with radius as 2. This could be caused by network partition and unusually long paths of reaching some nodes. When radius is 1, all nodes can be reached in one broadcast.

In Fig. 10, we compare our suggested search set with several related schemes: the incremental ERS scheme, California split, and the L-threshold scheme. The relative cost of our suggested search set can be seen to be lower than all other schemes. Note that search cost of our scheme is 2–10% higher than the optimum search set. This is because of the sparse network condition that invalidates (8).

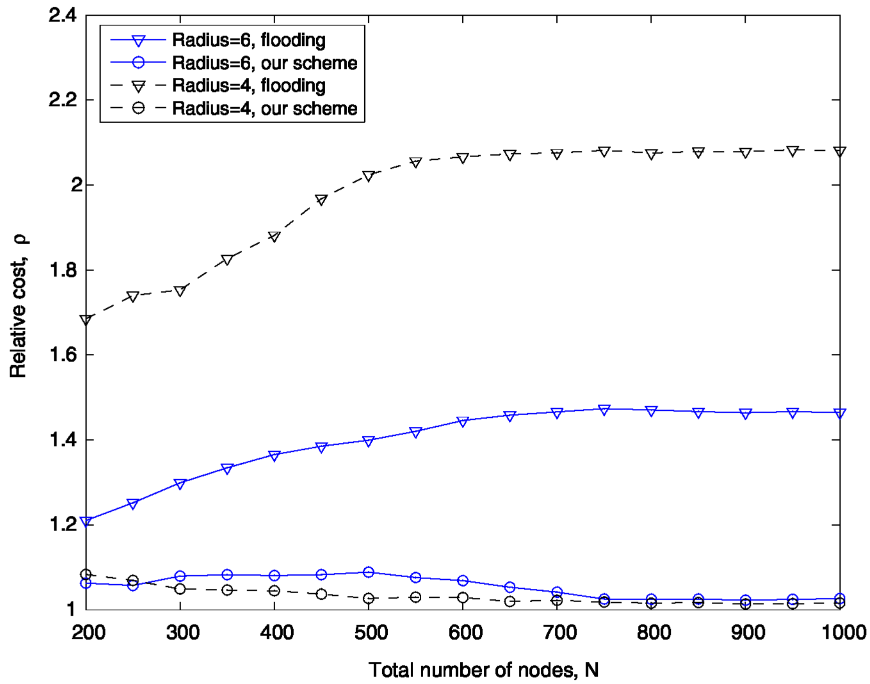

Fig. 8. Relative cost of our suggested search set and pure flooding scheme. The search costs have been normalized over the search cost of the optimum search sets, i.e., (19). The network radii are 4 and 6, respectively, and the source sits at the origin of the circle. The density of the networks can be calculated from N and the network radius. For example, when N is 200 and radius is 6, the density is $200/(\pi \cdot 6^2) = 1.76$ node per unit area.

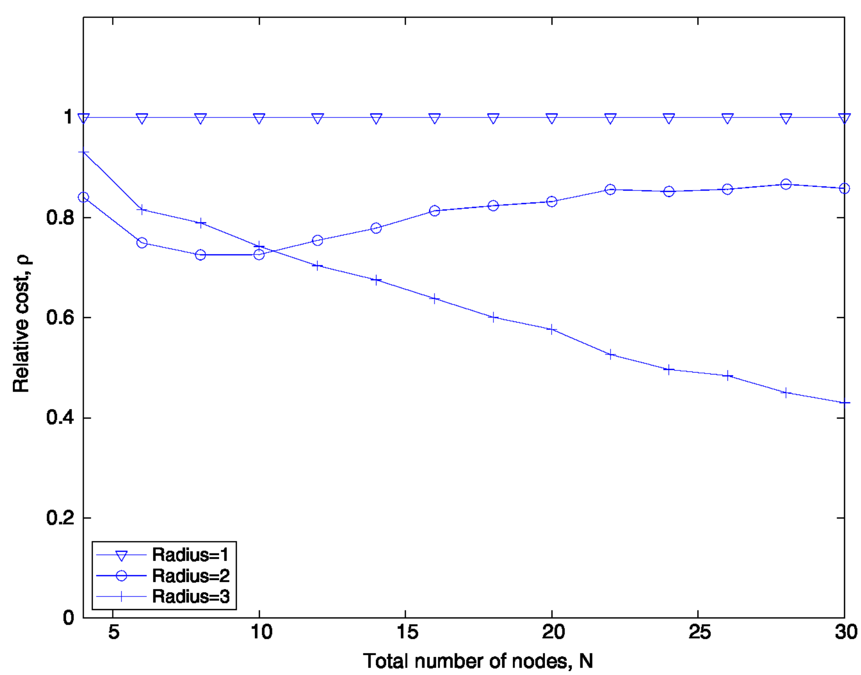

Fig. 9. Relative cost of our suggested search set and pure flooding scheme in sparse and small networks. We investigated networks with radius of 1, 2, and 3. The density of the networks can be calculated from N and the network radius. For example, when radius is 3, the density changes from 0.14 to 1.06 per unit area as N increases from 4 to 30.

6.2. Source randomly chosen from X

When the source may reside in any location within $X = b_R$, the optimum integer search sets differ from those shown in Table 1. Table 2 presents the optimum search sets corresponding to some relatively small R values. Interestingly, the maximum N in such optimum search sets is 4. This suggests that, even as R increases, the number of elements in the optimum search set is at most 4.

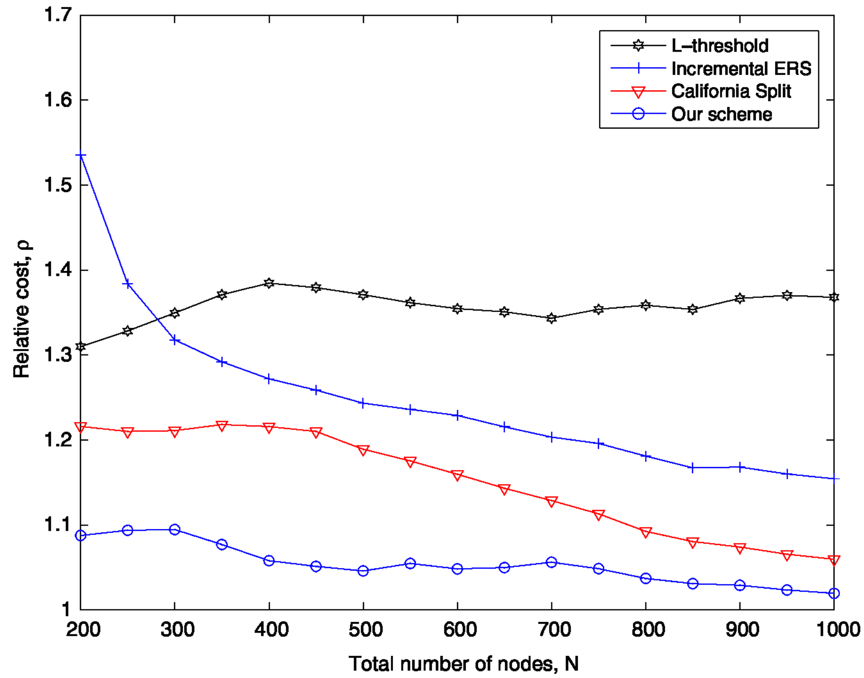

Fig. 10. Relative cost of our suggested search set, incremental ERS, California split, and the L-threshold scheme [8]. The search costs have been normalized over the search cost of the optimum search sets, i.e., (19). The network radius is 5 and the source sits at the origin of the circle. In our scheme, the search set is $\{r_1, r_2\} = \{\lfloor (R-1)/2 \rfloor, R-1\}$. The search set of incremental ERS is $\{1, 2, 3, \ldots, R-1\}$. California split uses a search set of $\{1, 2, 4, 8, \ldots\}$. The L-threshold scheme was suggested as $\{1, 2, 3, \ldots, L\}$ then flooding. The value of L was suggested to be 3 [8].

Table 2
Optimum integer search sets for different R

R	1	2	3	4	5	6	7	8	9	10	11	12
r_1^*	1	1	1	2	3	1	1	1	1	1	1	1
r_2^*		3	3	5	7	5	6	6	7	8	9	10
r_3^*			5	7	9	10	12	13	15	17	19	21
r_4^*						11	13	15	17	19	21	23

The source is assumed to be chosen randomly from the network. For example, the optimum search set is $r_1^* = 2$, $r_2^* = 5$, $r_3^* = 7$ when $R = 4$.

Fig. 11 presents the normalized search cost, p, of flooding and an $N = 2$ search set $\{r_1 = R; r_2 = 2R - 1\}$ over the minimum search cost in each R. It can be observed that the cost of flooding is significantly higher than that of the optimum search sets, even though this cost is decreasing as R increases. Instead of presenting the optimum search sets, we show the search cost of a search set with two elements, $N = 2$, with r_1 = R and $r_2 = 2R - 1$. Interestingly, we observe that such a simple search set introduces rather low extra cost over the optimum search cost. In fact, when $R > 13$, the extra cost of the $N = 2$ search set is at most 1% higher than the minimum search cost. Therefore, we recommend the use of this simple search set for the scenario where the source is chosen randomly from a circular network region with radius $R > 13$.

7. Conclusions and future works

Expanding ring search (ERS) is a widely- employed technique in computer networks especially in wireless networks in order to find services, destinations, or servers. In this work, we have developed a general analytical framework to study different search sets. Using this framework, we have been able to investigate the optimum search set when the source is at the origin of the circular network region. The optimum search set of such networks should be $\{r_1 ; r_2 \} = \{ L(R — 1)=2]; R — 1 \}$ with R as the radius of the circular network region. In the scenarios where the source is randomly chosen from the network, we have provided an almost-optimal search set, $\{r_1; r_2\} = \{R; 2R — 1 \}$, that guarantees the search cost to be at most 1% higher than the minimum search cost when R is relatively large, i.e., $R > 13$.

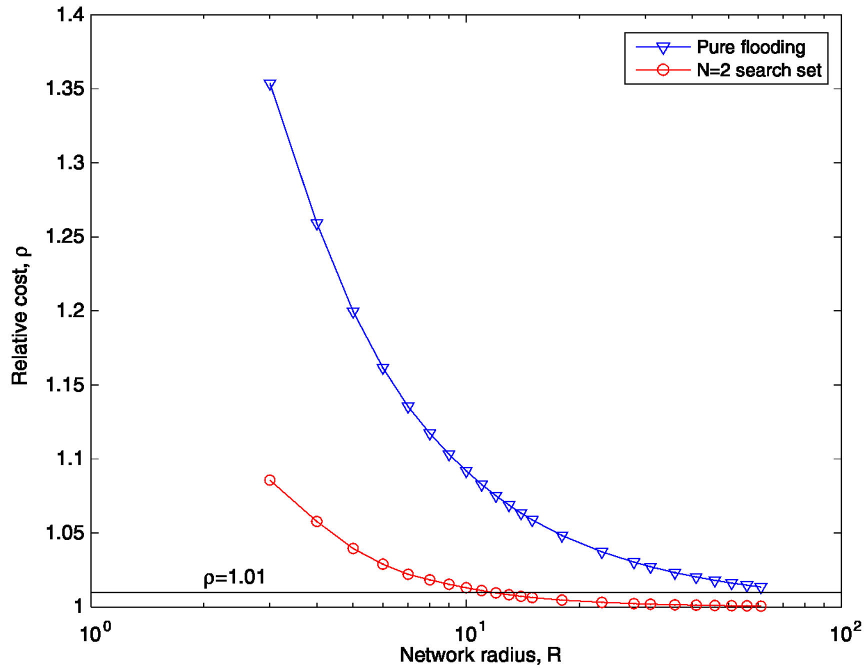

Fig. 11. Relative cost of pure flooding and $N = 2$ search set $\{r_1 = R, r_2 = 2R - 1\}$. The search costs have been normalized over the search cost of the optimum search sets, i.e., (19). Source is assumed to be randomly chosen from the network. Only integer search sets are allowed.

The main contribution of our work is that we have identified a search set that performs close to the optimum search set, which is hard to find. We have developed a general analytical framework to find such search sets. The framework can serve as a tool for further study in this field. We also provide guidelines for search set selections under various network setups.

We have assumed that the destination is chosen from the network randomly. In some applications, nodes in some regions are more likely to be chosen. How will such a destination selection probability distribution affect our optimum search set? In addition, it is intuitive that node density affects the network connectivity and topology. Will it change the optimum search set as well? Furthermore, our discussions have been limited to finding one destination. When multiple destinations need to be found simultaneously, what search sets should be used? Search delay would also be an interesting subject of study. For instance, how can we identify the best search set that satisfies a certain delay upper bound? We leave these interesting questions to our future work.

References:

1. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, IEEE Communications Magazine 40 (8) (2002) 102–114.
2. Y. Baryshnikov, E. Coffman, P. Jelenkovic, P. Momcilovic, D. Rubenstein, Flood search under the california split rule, Operations Research Letters 32 (3) (2004).
3. N. Chang, M. Liu, Revisiting the TTL-based controlled flooding search: optimality and randomization, in: Proceedings of the 10th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom’04), Philadelphia, PA, USA, September 26–October 1, 2004, pp. 85–99.
4. N. Chang, M. Liu, Controlled flooding search with delay constraints, in: Proceedings of the 25th Conference of the IEEE Communications Society (Infocom’06), Barcelona, Spain, April 23–29, 2006.
5. Z. Cheng, W.B. Heinzelman, Flooding strategy for target discovery in wireless networks, in: Proceedings of the ACM (MSWiM’03), San Diego, CA, USA, September 19, 2003.
6. J. Deng, Locating randomly selected destinations in large multi-hop wireless networks, in: Proceedings of the 19th International Teletraffic Congress (ITC-19), Beijing, PR China, August 29–September 2, 2005.

7. Z.J. Haas, J. Deng, B. Liang, P. Papadimitratos, S. Sajama, Wireless ad hoc networks, in: John G. Proakis (Ed.), Wiley Encyclopedia of Telecommunications, John Wiley & Sons, 2002.
8. J. Hassan, S. Jha, Optimising expanding ring search for multi-hop wireless networks, in: Proceedings of the IEEE Global Telecommunications Conference/Wireless Communications, Networks, and Systems (GLOBECOM'04), Dallas, TX, USA, November 29–December 3, 2004.
9. D.B. Johnson, D. Maltz, Y.-C. Hu, The dynamic source routing protocol for mobile ad hoc networks (DSR), April 2003, Internet Draft.
10. B. Krishnanmachari, Joon Ahn, Optimizing data replication for expanding ring-based queries in wireless sensor networks, in: Proceedings of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'06), Boston, MA, USA, April 2006, pp. 1– 10.
11. Q. Lv, P. Cao, E. Cohen, K. Li, S. Shenker, Search and replication in unstructured peer-to-peer networks, in: Proceedings of the ACM International Conference on Super- computing (ICS'02), New York, NY, USA, June 2002, pp. 84–95.
12. S. Ni, Y. Tseng, J. Sheu, The broadcast storm problem in a mobile ad hoc network, in: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99), Seattle, WA, USA, August 1999, pp. 152–162.
13. C.E. Perkins, Ad hoc on-demand distance vector (AODV) routing, February 2003, Internet Draft.

Notes:

1. We do not restrict these r_i values to integers in our framework. Integer search sets will be investigated in Section 6.
2. In fact, all optimum search sets were numerically calculated through the fmincon package in Matlab. Basically, we looked for an optimum N and t_1; t_2; ... ; t_{N-1} that minimize function (10) under the constraints of (11). These optimum search sets serve as the performance upper bound for all schemes.
3. http://www.R-project.org.
4. Note the different notations of the search sets used in this paper and [6].