A Query Scope Agent for Flood Search Routing Protocols

JOHN SUCEC * and IVAN MARSIC
Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854-8058, USA

Abstract. Flood-search on-demand routing has received considerable interest for its application to mobile ad hoc networks. To alleviate the effects of flooding the network with control packets to discover a route, the concept of an expanding ring search (ERS) has been proposed elsewhere for reducing the packet transmission overhead of the route discovery process. Essentially, ERS consists of incrementally increasing the allowable hop radius of the flood search until a route to the target node is returned. However, ERS incurs additional latency to successfully complete the route discovery procedure. This paper presents a query scope agent (QSA) that assists in the selection of an appropriate ERS. The QSA accepts as input, from the user or network application, a maximum allowable value for route discovery delay. The QSA then estimates network parameter values to determine an ERS approach that satisfies the delay requirement while reducing expected packet transmission overhead. Simulation results show that it successfully achieves this objective. Further, the QSA incurs little communication and computation overhead, and operates in a distributed and asynchronous fashion.

Keywords: mobile ad hoc network, on-demand routing, flood search

1. Introduction

Network protocols based on reactive routing techniques have received considerable attention due to their applicability to mobile ad hoc networks (MANETs). Among the routing protocols that are based at least in part on reactive routing are the Ad hoc On-Demand Distance Vector (AODV) routing protocol [15,16] and the Dynamic Source Routing (DSR) protocol [7,8]. These particular protocols are reactive, or "on-demand", in nature because each proactively acquires little or no data regarding the network topology. Instead, when a node originates a packet to a destination for which no forwarding path is known, the originating node initiates a *flood-search* route discovery procedure.

Ignoring the implementation details of any specific flood-search on-demand routing protocol, a generalized description of the route discovery procedure is as follows. The procedure consists of an originating node, or source node (s), disseminating a route request (RREQ) packet that effectively queries recipient nodes for a route to the target node (t). (Assuming a shared media, broadcast transmission environment, the RREQ packet can be propagated by a node to each of its neighbors via a single broadcast transmission.) If the RREQ arrives either at t or at a node with a valid path to t in its route cache, a route reply (RREP) message is sent back to s. The RREP packet is returned along the reverse of the path traversed by the received RREQ packet until it arrives at s.

1.1. Query scope selection

The issue of primary interest in this paper is setting the query radius (R_Q), measured in hops from s, to which the RREQ should be propagated. At most, R_Q may need to be set to D, the network diameter. Under this condition, the route discovery procedure is equivalent to network-wide flooding (NWF) and may result in up to $|V| - 1$ transmissions of the RREQ packet.[1] That is, every node except t itself propagates the packet. At the optimistic end of the spectrum, s may presume that with high likelihood a neighbor of itself has a route to t in its cache. If this is indeed the case, then a single transmission of the RREQ packet by s is sufficient. DSR specifies such a non-propagating query in its implementation and resorts to NWF if a RREP is not received by s within some maximum allowable non-propagating query duration [8].

In addition to setting R_Q to either 1 hop (a non-propagating or local query) or D hops (a NWF query), R_Q may be set to anything in between. A number of heuristics can be employed to determine R_Q. Among them include the expanding ring approach as detailed in [16] for application to AODV. In the expanding ring approach, if a timer for the existing query expires, R_Q is incremented by some amount and a new query is initialized. To be consistent with terminology of [16] and [8], the process of incrementally increasing the query radius and allowable query duration until the query scope reaches either t or a node with a route to t in its route cache, is referred to here as an *expanding ring search* (ERS).

The benefit of employing ERS is that the expected number of RREQ packet transmissions (ψ) may be reduced from what would be incurred if a network-wide flood search is applied immediately for each route discovery event. Although an ERS procedure that defers NWF for as long as possible may minimize the ψ metric under certain network conditions, it may also incur an unacceptable expected route discovery delay (τ). This occurs when each incremental increase in R_Q yields little additional likelihood of route discovery over the previous

* Corresponding author.
  E-mail: jsucec@caip.rutgers.edu

---

[1] Where, V is the set of nodes in a given network, E is the set of bi-directional links connecting nodes in V and $G = (V, E)$ is the undirected graph formed by V and E. $|S| \equiv$ cardinality of the set S.

value of R_Q. Under such circumstances, it is desirable for the route discovery procedure to employ a larger R_Q increment or resort immediately to NWF.

1.2. Example of ERS performance

To demonstrate the benefit of ERS for route discovery, ψ and delay are compared for several ERS approaches in a 100-node network. The results are depicted in figures 1 and 2. The ERS approaches under consideration are as follows:

- $R_Q \in \langle D \rangle$, i.e., flood immediately (*0-ERS*),
- $R_Q \in \langle 1, D \rangle$, i.e., $0 \to 1$ hop (*1-ERS*),
- $R_Q \in \langle 1, 2, D \rangle$, i.e., $0 \to 1 \to 2$ hops (*1-2-ERS*),
- $R_Q \in \langle 1, 2, 4, D \rangle$, i.e., $0 \to 1 \to 2 \to 4$ hops (*1-2-4-ERS*).

For each ERS approach, a simulation consisting of 100 trials was performed. In each simulation, $R_{TX} = 250$ m and the
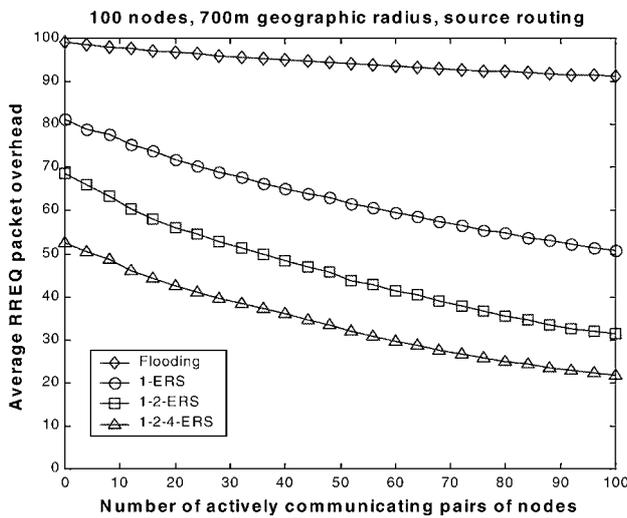
Figure 1. ERS comparison, packet overhead.
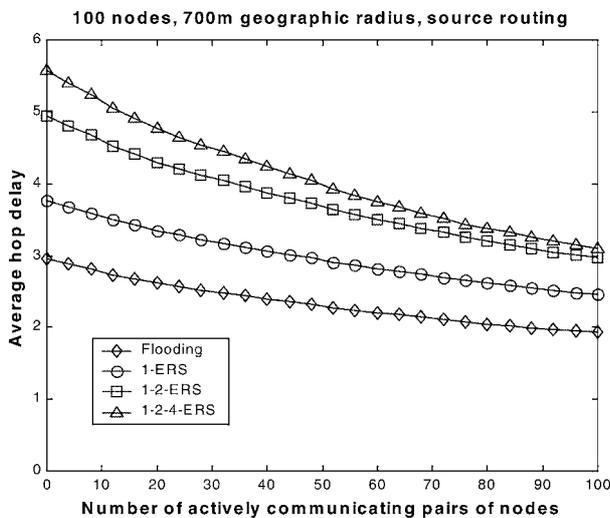
Figure 2. ERS comparison, delay (in hops).

network area is a circle with a radius of 700 m. All other network conditions are as detailed in sections 2.1 and 4.1.

To simplify the assessment of route acquisition latency (τ), the concept of *hop delay* (η) is employed for the simulations herein. Hop delay is defined as the total number of hops that must be searched in the route discovery procedure before a RREP is transmitted back to s. As an example, a case is considered where the initial RREP is generated by a node situated 5 hops away from s. If *1-2-ERS* happens to be in effect, $\eta = 8$ hops: 5 (hops searched during NWF) + 1 (due to failed 1-hop query) + 2 (due to failed 2-hop query) = 8 hops. Presuming τ is proportional to η, due to the linear dependence of the duration for a route query timer on the hop radius of a route query, assessing route acquisition latency in terms of hop delay is a reasonable simplification. This simplification is revisited in Section 5.

As indicated by figure 1, the most gradual form of ERS (in this case *1-2-4-ERS*) has the lowest value for ψ. This is because, for the scenario under consideration, gradually incrementing the query radius successfully avoids a costly network-wide flood search with high likelihood. On the other hand, *1-2-4-ERS* also incurs the highest value for η, as indicated by figure 2. This is because whenever a query failure occurs, whether R_Q is 1, 2 or 4 hops, a timer must expire before *1-2-4-ERS* increments R_Q to the next value (either 2, 4 or D hops). Also, as the number of *pairs* of communicating nodes (P) increases, the performance of all ERS approaches improves in terms of ψ and η due to the increased amount of routing information that may be acquired via route tapping.

As evidenced by figures 1 and 2, a wide range for ERS performance is possible. Hence, the motivation for a QSA that can judiciously decide whether to employ gradual ERS or coarse-grained ERS based on an application's allowable route discovery latency. As a final note regarding figures 1 and 2, the results shown correspond to the source routing case. Results for the next-hop routing case are similar although the curves for next-hop routing tend to be less steeply sloped.

1.3. Paper outline

This paper presents a novel procedure, known hereafter as a *query scope agent* (QSA), for selecting the type of ERS to be employed. The purpose of the QSA is to select an ERS approach that reduces ψ while keeping route discovery delay less than some maximum allowable value specified by a user or application.

The rest of the paper is organized as follows. Section 2 presents the framework upon which the simulation work described herein is based and briefly summarizes some related work. Section 3 describes the QSA implementation. Section 4 reports simulation results and assesses the performance of the QSA. Lastly, section 5 summarizes the key points of this paper.

2. Framework

2.1. Network environment

The underlying network and link assumptions for this paper are as follows. The network topology is represented by a *connected* and *undirected* graph $G = (V, E)$. Every node is equipped with a transceiver whose transmission range is given by R_{TX}, in meters. All nodes within the transmission range of a node v will be able to hear transmissions by v. Likewise, v will be able to hear the transmissions of all nodes lying not more than R_{TX} from it. On the other hand, nodes lying more than R_{TX} from one another are assumed to be unable to communicate directly with one another and at least one intermediary node is required to forward packets between such pairs. The MAC protocol on all interfaces is some variant of CSMA/CA. It is assumed that receivers operating in promiscuous mode can tap the frames transmitted on the CSMA/CA media.

The network layer addressing protocol can be any protocol that supports both next-hop routing and source routing (e.g., IPv6). Additionally, the network layer datagram header must provide a means to indicate the path length between s and t. This can be provided implicitly when source routing is in effect. When next-hop routing is in effect, a header extension or an extra field is needed in the datagram header to indicate explicitly the path length. Transmission of path length information is needed to estimate the number of pairs of communicating nodes, as discussed in section 3.2.

Each network node has a single CSMA/CA network interface card (NIC) and, therefore, each node can be uniquely identified by the address associated with that NIC. Further, it is assumed that each node is cognizant of $|V|$ and knows the node ID (i.e., the network address) of all network nodes. This information is possibly pre-configured in network nodes. If a new node joins the network then a single NWF is performed for it to announce itself. (The new node would also have to query a neighbor to obtain node count and node ID information of the network nodes.) Presumably, the frequency of join (or departure) events will be sufficiently low so that overhead associated with tracking network node count and node identities does not contribute significantly to network traffic. Lastly, it is assumed that the routing protocol includes a variant of the Hello protocol that allows each node to periodically announce itself to its neighbors and, therefore, discover its neighbors.

2.2. Routing paradigms

Although AODV and DSR are cited in this paper, the specifics of their implementations are not essential for the purposes here. Instead, AODV and DSR are referenced primarily as representatives of two important classes of routing protocols: next-hop table-lookup routing and source routing, respectively. A third class of routing protocols is geographic routing, which has received attention recently [9,11]. The geography-based routing paradigm will not be considered here.

The reason why next-hop routing and source routing deserve separate consideration in terms of route discovery is because their potentials for passively acquiring topology information via tapping of incident packets can be quite different. Whereas a next-hop routed packet contains only the source and destination addresses in the datagram header, a source routed packet contains also the addresses of the intermediary nodes of the source to destination path specified in the datagram header. Thus, when nodes operate their NICs in promiscuous mode, as assumed in the simulation results of [1,4,12], a node tapping the network layer headers of packets transmitted by its neighbors will potentially learn significantly more topology information if the network employs a source routing protocol than it would if a next-hop routing protocol was employed. Since such passively acquired topology data can be used in the route discovery process, source routing protocols have a greater potential for efficient route discovery than next-hop routing protocols. Of course, route acquisition efficiency is only one possible measure of routing protocol performance.

2.3. Earlier work

A considerable number of simulation studies have been conducted to assess the performance of various MANET routing protocols that have on-demand characteristics. Among these include [1,3,5,6]. Further, algorithms and heuristics have been proposed for alleviating the effect of network-wide flooding in MANETs. Among these include [13,14,18]. However, none of this earlier work has assessed the performance of an ERS-like approach to reducing routing protocol overhead or considered distributed parameter estimation techniques to predict actual network conditions.

Ko and Vaidya [10] propose a gradually expanding request zone in order to reduce the likelihood of NWF events. Unlike [10], however, the methods of this paper do not presuppose the presence of GPS enabled nodes or the availability of GPS data in the route discovery process.

Heuristics for query containment that exploit knowledge of a previously known valid path to a target node t are proposed in [2]. The underlying premise of these techniques is that many of the nodes lying on a previously valid path are likely to be still useful for constructing a new path to t and yields a localization of the query process when node mobility has not drastically disrupted the earlier known topology. This approach appears to be very promising for the purpose of *route maintenance*. However, in cases where no path to t is previously known or when path information becomes stale as a result of node mobility, such query localization techniques are ineffective. The methods of this paper, on the other hand, do not presuppose knowledge of topology data specifically related to finding a route to a particular target node. Thus, this paper addresses query containment issues beyond the scope of [2].

Perkins et al. [16] describe how ERS may be applied in the context of AODV routing. When performing route discovery for a target that was previously queried, AODV sets R_Q initially to the hop distance set in the RREP packet of the most

recent query to that node, plus a query increment constant. Presuming the hop distance at which a RREP for the previous query was generated is still relevant to the current network conditions, then this may serve as a reasonable initial value for R_Q. However, if there is no previous query to t, or network mobility has rendered the hop distance of the previous RREP obsolete, then the method of [16] for initializing R_Q is not an effective means for controlling ERS. Further, if the initial query fails, Perkins et al. [16] simply specify that R_Q be incremented by a predetermined value and does not provide a dynamic means by which to increment R_Q that is responsive to network conditions.

In [17], procedures for a routing protocol to decide between *1-ERS* and *1-2-ERS* are proposed. Although the methods of [17] are responsive to the prevailing network conditions and facilitate an intelligent trade-off between ψ and τ, they do not permit a user or application to specify a ceiling for τ. The QSA proposed here, therefore, provides additional flexibility. Further, the QSA includes enhancements to support more robust performance and lower computation overhead.

3. Query scope agent (QSA)

The rationale underlying the QSA proposed here is that if the probability mass function (PMF) $f_H(h)$ for the hop distance (H) at which a route to t is cached can be predicted, then an appropriate ERS approach can be selected. Considering figures 3 and 4 as examples of hop count PMFs, it is intuitively evident that the PMF of figure 3 will yield savings in terms of ψ when a gradual form of ERS is applied. The PMF of figure 4, on the other hand, will *not* yield significant savings in terms of ψ irrespective of the ERS. For a network environment corresponding to figure 4, therefore, it may be preferable to simply initiate a network-wide flood search rather than incrementally increase the scope of the search and possibly incur excessive route acquisition delay. If the QSA can discern whether the network conditions correspond to a PMF conducive to one type of ERS versus another then ψ can be reduced (versus immediate network-wide flood search) *without* substantially increasing τ.

3.1. QSA overview

To describe the QSA, the following definitions are useful.

- $\deg(v) \equiv$ the degree of node v, i.e., the number of neighbors of v;
- $N(v) \equiv$ set of nodes in the *open* neighborhood of v;
- $H \equiv$ random variable corresponding to the shortest distance, in terms of hops, from a source node s at which routing information is available for a target node t;
- $\eta \equiv$ expected hop delay for route acquisition ($\eta \propto \tau$);
- $\eta_{max} \equiv$ maximum value of η allowable for a given application;
- $P \equiv$ number of currently active communication sessions in the network;
- $L \equiv$ average communication path length (in hops);
- $N_k \equiv$ set of nodes in the *closed* k-hop neighborhood (of v);
- $\Pi \equiv$ the set of paths known to v;
- $\Pi_k \equiv$ a path belonging to the set Π where $k \in \{1, 2, \ldots, |\Pi|\}$;
- $p_{j,k} \equiv$ incremental ERS success probability (i.e., the probability of route acquisition success when the flood-search scope is incremented from j hops to k hops).

Figure 5 depicts the software interfaces for the QSA and its relation to some of the other entities that impact routing. The essential functionality of the QSA is to select an ERS approach, when triggered by a route query event, that satisfies the maximum allowable hop delay (η_{max}) required by an application while minimizing the expected number of RREQ packet transmissions (ψ).

To achieve the objective of reducing ψ while satisfying the η_{max} requirement, the QSA accepts as input from the routing protocol (e.g., AODV or DSR), path information (Π). Further, the QSA running at node v learns from a Hello protocol
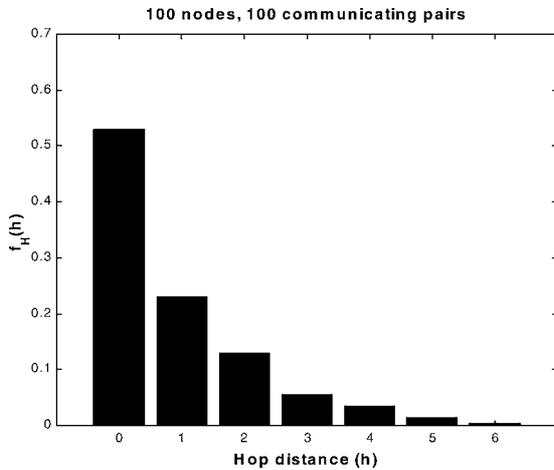
Figure 3. PMF conducive to gradual ERS.
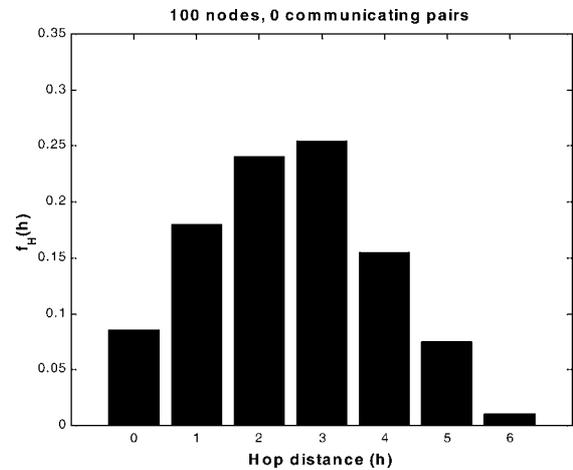
Figure 4. PMF conducive to coarse-grained ERS.
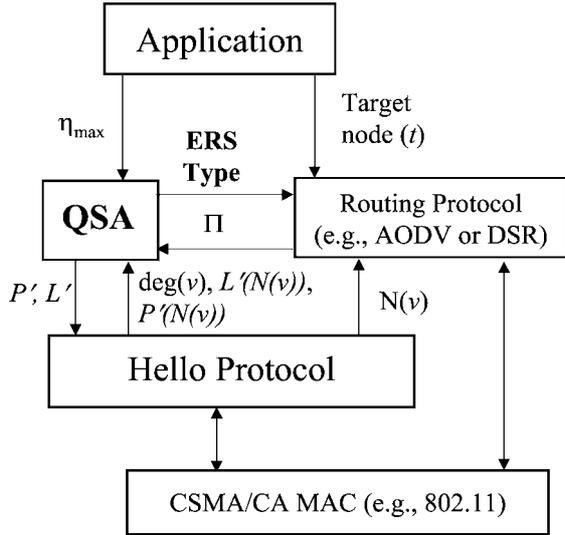
Figure 5. Software interfaces for the QSA.

Figure 6. Simulated node placement.
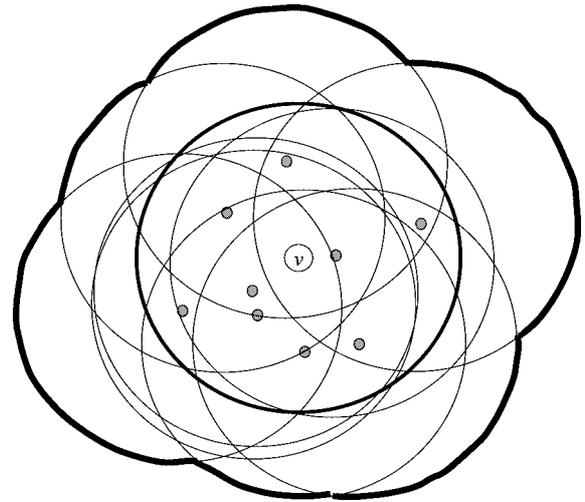
Figure 7. 2-hop region of coverage.

$\deg(v)$. The QSA then uses Π and $\deg(v)$ to compute the estimates of P, L and $|N_k|$.

The estimates of P and L are shared with the neighbors of v via the Hello protocol. Likewise, nodes in $N(v)$ share their own estimates, $P'(N(v))$ and $L'(N(v))$, via the Hello protocol. Lastly, the QSA at v employs its neighbors' estimates to compute refined estimates $P' \rightarrow P''$ and $L' \rightarrow L''$.

Knowledge of $\deg(v)$ is used to estimate $|N_k|$, $k \in \{1, 2, \ldots, D\}$. Combining the estimates of P, L and $|N_k|$, the QSA estimates the probabilities for successful route discovery due to incrementing the query scope. These probability estimates are then used to estimate η. The estimate, η', is then compared with $\eta_{\max}$ to obtain a set of feasible ERS approaches. The most *gradual* ERS approach in this set is selected as the ERS strategy to be employed for the route discovery procedure. A summary of the QSA routine is as follows:

1. Estimate network parameters:

   - number of pairs of communicating nodes (P),
   - mean communication path length (L),
   - node count of k-hop neighborhoods ($|N_k|$).

2. Estimate ERS success probabilities, $p_{j,k}$.

3. Estimate η for various ERS approaches.

4. Select best (predicted) ERS method.

Estimation of P, L and $|N_k|$ is necessary because of the intrinsic dependence of $f_H(h)$ on these parameters. The estimates are then used to predict $p_{j,k}$ and η. The details of this routine are provided in sections 3.2–3.4.

3.2. QSA parameter estimation

The first step of the QSA procedure for selecting an appropriate ERS approach is to estimate the number of active communication sessions (P) and the average path length for these sessions (L). From the routing protocol, the QSA receives information concerning the set of active paths that have been learned via route tapping or route snooping (Π). From the Hello protocol, the QSA learns information regarding the one-hop neighborhood. Estimates for P and L are then computed as follows:

$$P' = \frac{|V|}{1 + \deg(v)} \sum_{k=1}^{|\Pi|} \frac{|N_1 \cap \Pi_k|}{|\Pi_k|}, \qquad (1)$$

$$L' = \frac{1}{|\Pi|} \sum_{k=1}^{|\Pi|} |\Pi_k|. \qquad (2)$$

Averaging over the estimates computed by neighboring nodes to yield P'' and L'' subsequently refines the values outputted from (1) and (2).

Next, the QSA estimates the cardinality of the k-hop neighborhoods about v. The 1-hop neighborhood is provided by the Hello protocol. Estimation of $|N_k|$, $k > 1$, is more complex. The procedure is recursive and most easily explained by considering first the estimation of $|N_2|$.

Using the $\deg(v)$ information provided by the Hello protocol, the QSA simulates random placement of $\deg(v)$ nodes within R_{TX} of v. Figure 6 shows a possible arrangement of randomly situated nodes.

$R_2 \approx 1.6 \times R_{TX}$
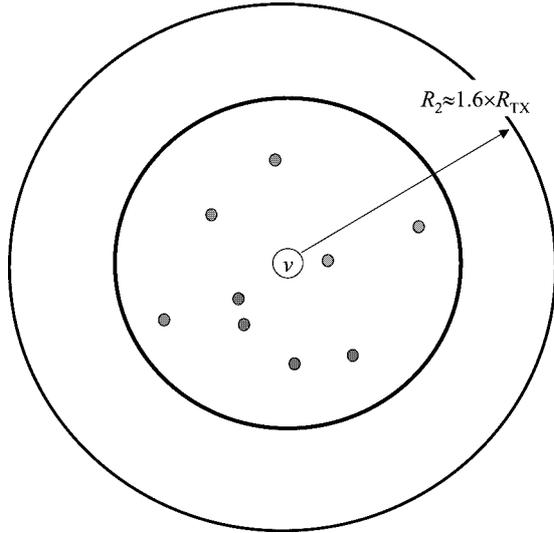
Figure 8. 2-hop circular area.
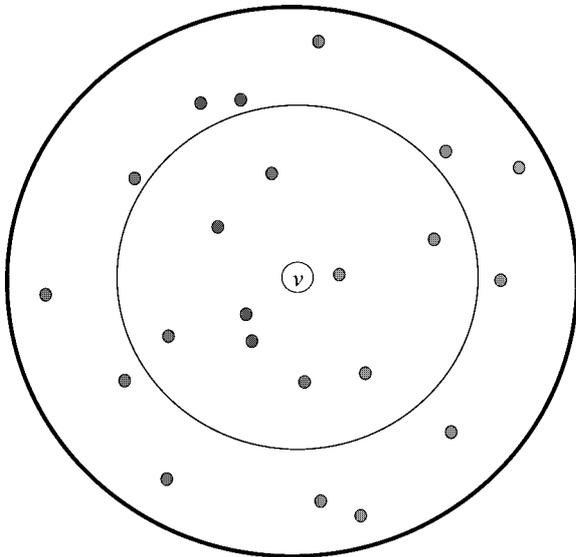
Figure 9. 2-hop simulated node placement.

Upon generation of simulated node placement, as in figure 6, the QSA then computes the area of coverage provided by these nodes. Figure 7 shows the actual region covered by the transmissions of nodes in N(v) while figure 8 shows the circular region of approximately equal area. Letting $\delta = |N_1|$, the number of nodes number of nodes in N_2 is estimated as follows:

$$|N_2|' = \left(\frac{R_2}{R_{TX}} \right)^2 \cdot \delta. \tag{3a}$$

The R_2 term of (3a) corresponds to the estimate of the 2-hop neighborhood geographic radius, as illustrated in figure 8.

The procedure used to estimate $|N_2|$ is applied in a recursive manner to estimate $|N_k|$, $k > 2$. Figure 9, for example, shows a possible randomly simulated 2-hop node distribution after estimating $|N_2|$.

After estimating R_k by computing the area covered by a transmitting set of $|N_{k-1}|'$ nodes, $|N_k|'$ is given as follows:

$$|N_k|' = \left(\frac{R_k}{R_{TX}} \right)^2 \cdot \delta. \tag{3b}$$

The recursive procedure for estimating $|N_k|$ may stop when $|N_k|' \geqslant |V|$ or for some suitably large value of k, afterwards $|N_{k+j}|$ is estimated based on a linear incremental growth in neighborhood size. That is, $|N_{k+j}|' = |N_k|' + jm$, for some integer constant m.

An important detail implicit in the preceding procedure for estimating the k-hop neighborhood size, is that it has been assumed that v is situated at the geographic center of the network area. Applying such an assumption yields estimates of $|N_k|$ that tend to be too high. In the appendix, a more robust procedure is described and has been employed for the simulations reported in section 4. This procedure accounts for the fact that v will be typically offset from the geographic center of the network and that the k-hop geographical area of transmission coverage may in fact be "wasted" on regions that are outside the geographical area of the network.

Lastly, the estimates of $|N_k|$ are used to compute a rough estimate of E[$H \mid H \geqslant 1, \ P = 0$]:

$$\text{E}[H \mid H \geqslant 1, \ P = 0]' = \frac{\sum_{h=1}^{D-1} h \cdot (|N_{h+1}|' - |N_h|')}{|V| - |N_1|'}. \tag{4}$$

3.3. ERS success probabilities

In this section, estimates of ERS success probabilities ($p_{j,k}$) are described. The estimates are based on knowledge of $|V|$ and estimates of the parameters P, L and $|N_k|$. Once the $p_{j,k}$ terms are estimated, the expected hop delay (η) for the relevant ERS approaches is estimated.

Defining X as the set of nodes transmitting datagrams with the IP address of node t in its header, it is clear that the larger $|X|$ is, the more likely that a route to t will be cached at an arbitrary node. Thus, a step that remains in estimating $p_{j,k}$ is to estimate $|X|$. Estimates of $|X|$ for the source routing and next-hop routing cases are given as follows:

$$|X_{source}|' = \frac{(P'')^{0.9} \cdot (L'')^2}{|V|}, \tag{5a}$$

$$|X_{next-hop}|' = \frac{(P'')^{0.9} \cdot (2 \cdot L'')}{|V|}. \tag{5b}$$

The selection of the exponent for P'' in (5) is somewhat arbitrary. The exponent value given was selected to reflect, to some degree, the effect of *overlapping* communication paths. That is, just as a target node t may serve as a packet forwarder on multiple communication paths, many of its peers are likewise packet forwarders on more than one of these paths. Thus, although t may lie on multiple communication paths, $|X|$ typically grows sub-linearly in the number of paths on which t lies. $|X|$ depends also on L: As the average

path length increases, there will be more nodes on average involved with forwarding datagrams that contain the address of t in their header. In the case of source routing, the address of t appears in the header of all unicast datagrams that are forwarded by t. Thus, $|X|$ is quadratic in L for the source routing case, as per (5a). For next-hop routing, the address of t appears in the header only if t is the source or destination of the datagram. Thus, a factor of $2L$ is contributed in (5b). Lastly, $|V|$ appears in the denominator of (5) as t is only one of $|V|$ nodes.

An estimate of incremental ERS route acquisition probability is generated as follows. First an estimate of the *fraction* of nodes lying outside the scope of the j-hop radius of s (i.e., $v \in V - N_{j+1}$) with routing information for t is computed:

$$f'_{X|j} = \max\left\{ \frac{1}{|V| - |N_{j+1}|'}, \frac{|X|'}{|V|} \right\}. \tag{6}$$

Next, it is noted that successful route discovery occurs when the set of nodes lying in the set $N_{k+1} - N_{j+1}$ intersects X. In a 2-dimensional network, the nodes in $N_{k+1} - N_{j+1}$ essentially cover an annular region with inner radius of $j + 1$ hops and outer radius of $k + 1$ hops. The fraction of nodes lying in the range of $j + 1$ to $k + 1$ hops from s (and not previously queried) is estimated as follows:

$$f'_{j,k} = \frac{|N_{k+1}|' - |N_{j+1}|'}{|V| - |N_{j+1}|'}. \tag{7}$$

The communication paths on which t lies consist of nodes X that tend to be clustered about t (because such paths must contain t, itself), and therefore, X effectively covers some connected region of the network area. This effect of overlapping areas is illustrated in figure 10. Thus, to estimate the probability of intersection between $N_{k+1} - N_{j+1}$ and X, the technique employed here actually estimates the probability that the network region covered by $N_{k+1} - N_{j+1}$ overlaps the network region covered by X. To approximate the effect of *overlapping areas*, and thereby $p_{j,k}$, the principle of overlapping intervals on a unit line segment is applied, as in figure 10.

Applying this line interval model, $p_{j,k}$ can be efficiently estimated:

$$p'_{j,k} = \frac{\sqrt{f'_{X|j}}}{1 - \sqrt{f'_{j,k}}}. \tag{8}$$

Square roots appear in (8) in order to account for the fact that the Euclidean distance separating the centers of a pair of geographic regions in a 2-dimensional network is *square root in the overall network area*, on average. Lastly, it is noted that the subscripts given in (6) and (7) for the neighborhood sets are $j + 1$ and $k + 1$, to reflect the effect of the Hello protocol that allows nodes to learn their neighbors.

3.4. ERS selection

Once $p_{j,k}$ are estimated, they are used along with the estimate of $E[H \mid H \geqslant 1, P = 0]$ to compute estimates of η.

$$\eta'_{1\text{-}ERS} = 1 - p'_{0,1} + E[H \mid H \geqslant 1, P = 0]', \tag{9a}$$

$$\eta'_{1\text{-}2\text{-}ERS} = \eta'_{1\text{-}ERS} + 2 \cdot (1 - p'_{1,2}), \tag{9b}$$

$$\eta'_{1\text{-}2\text{-}4\text{-}ERS} = \eta'_{1\text{-}2\text{-}ERS} + 4 \cdot (1 - p'_{2,4}). \tag{9c}$$

Each of the predictions computed via (9) are compared with $\eta_{\max}$. If a predicted hop delay is less than $\eta_{\max}$ then the associated ERS method is considered feasible. The ERS that is most likely to minimize ψ is then selected, according to the following criteria:

- If *1-2-4-ERS* is in feasible set → use *1-2-4-ERS*;
- Else if *1-2-ERS* is feasible → use *1-2-ERS*;
- Else if *1-ERS* is feasible → use *1-ERS*;
- Else flood immediately (*0-ERS*).

The chosen ERS method is then passed to the routing protocol (e.g., AODV or DSR) for route discovery.

As a final note on the QSA implementation, nearly all of its computation is done off-line. That is, when a route request arrives with its associated $\eta_{\max}$, the QSA merely must compare the current η' with $\eta_{\max}$ and then select the ERS approach

Figure 10. Approximation of overlapping area by considering overlapping intervals on a line segment.

within the feasible set that is likely to yield the lowest ψ. In fact, much of the QSA is implemented very efficiently via closed form expressions. In terms of computational cost, the most expensive portion of the QSA is the calculation of the area of transmission coverage achieved by randomly situated nodes as performed for figures 7–9. The current implementation of this calculation by the QSA consists of a naive approximation method that is cubic in the node count. While more efficient techniques may exist, they have not yet been explored. Again, however, this calculation is performed non-real time during an interval *between* route request events, and only when the Hello protocol indicates that a change to $|N_1|$ has occurred. Thus, the actual deployment of the QSA incurs *little* additional route acquisition overhead (i.e., the time required to compute (5) and (9) and compare the result with the $\eta_{\max}$ threshold).

3.5. Comments on network assumptions

The QSA, as described here, considers networks where nodes are situated in a circular area. This is because the circular region represents the most pessimistic geometry for passively acquiring routing information via route tapping. That is, the circular geometry minimizes $E[L]$ and, therefore, also minimizes $E[|X|]$ for a given P and $|V|$, as per (5).

Of course, a circular network region is not the only possible geometry. Network nodes, for example, may be situated within a narrow rectangular strip or even in a snake-like or linear arrangement. Under such network geometry, the methodology employed by the QSA to estimate $|N_k|$ is sub-optimal. To account for this, extensions to the QSA may be employed. Considering circular geometry, $L = \Theta(\sqrt{|V|})$, but if there are numerous paths whose lengths are significantly longer than $\sqrt{|V|}$, the QSA should assume a rectangular network geometry. The width and length of the rectangular are predicted based on $|\Pi_k|$, $k \in |\Pi|$, and the estimate of L. Detail of how such an extension might be implemented is beyond the scope of this paper.

This paper has assumed nodes to be randomly situated in accordance with a 2-dimensional uniform distribution. In practice, though, nodes may tend to clump together into numerous cliques randomly situated throughout the network area. However, because the network is assumed *connected*, it is unlikely that such grouping will adversely impact the performance of the QSA. That is, since V is connected a subset $C \subset V$ corresponding to the set of cliques is also connected. Given this and the fact that the QSA uses a refined estimate of P and L based on estimates exchanged with neighbors via the Hello protocol, the QSA ERS selection should be robust in a clumped environment, as well.

4. Simulation results

Monte Carlo simulations, in accordance with the conditions described in sections 2.1 and 4.1 were conducted for cases of 50, 100 and 200 nodes. Only the results for the source routing

environment are reported here and are given in figures 11–16. The results for the next-hop routing environment are similar and are omitted here.

4.1. Simulation assumptions

For the simulations that generated the results of figures 11–16, network nodes are situated randomly throughout a network area of fixed size in accordance with a 2-dimensional uniform probability distribution. To assess the probability of route discovery, it is assumed that the network topology remains fixed over the duration of any given route discovery procedure. This topology *snapshot* assumption is valid under the condition where node mobility is sufficiently modest such that the network link state represented by E is unlikely to change over the duration of a single route discovery event. This is a reasonable assumption to make. Otherwise, if the ratio of node speed to R_{TX} while route discovery is taking place is high enough to cause significant changes to E, then any routing information conveyed by a RREP is likely to be
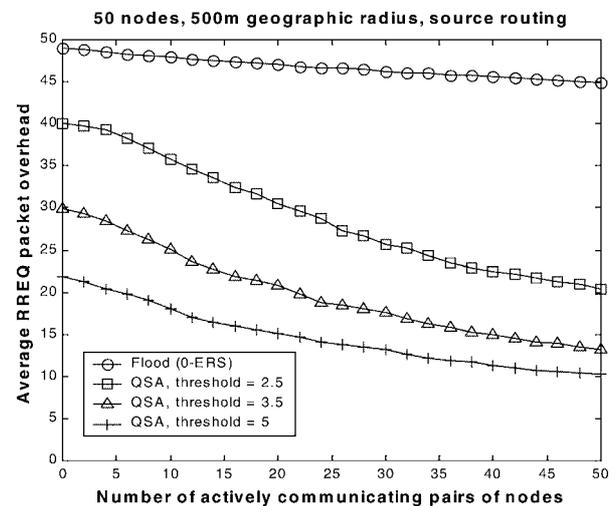
Figure 11. 50 nodes, RREQ packet overhead.
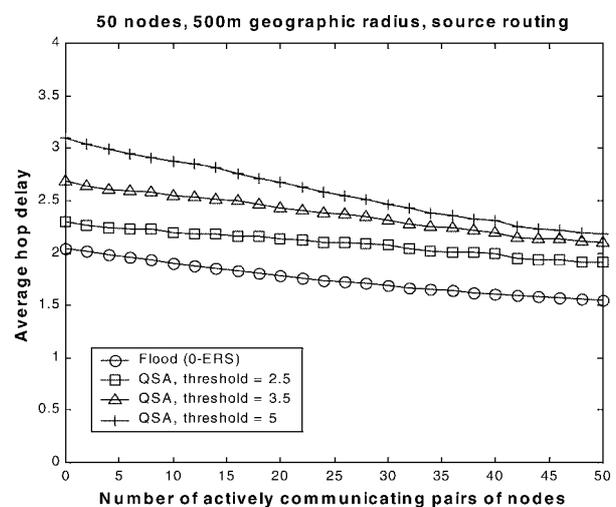
Figure 12. 50 nodes, hop delay.
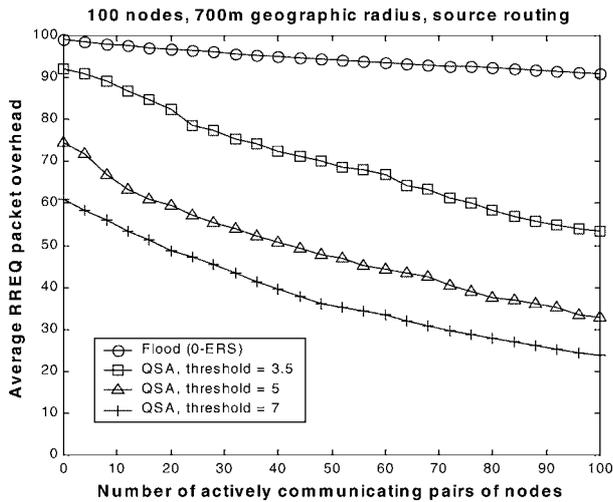
Figure 13. 100 nodes, RREQ packet overhead.
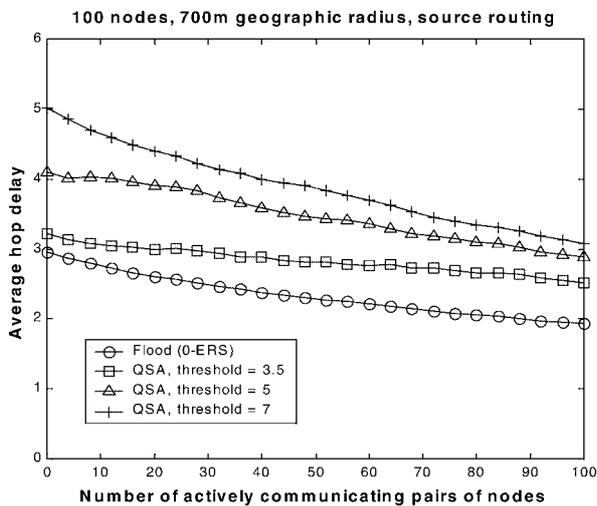
Figure 14. 100 nodes, hop delay.
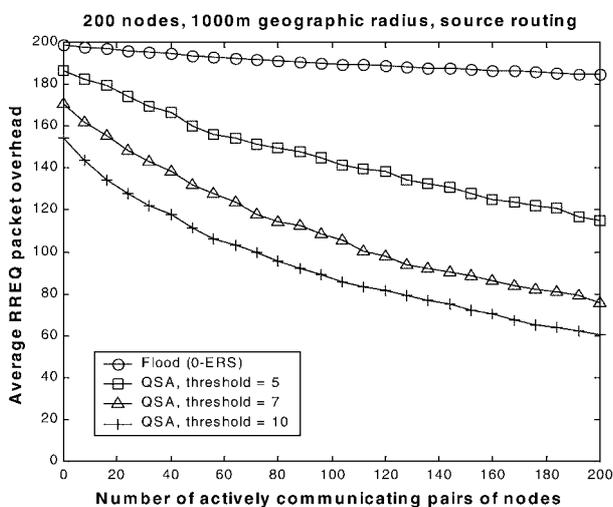
Figure 15. 200 nodes, RREQ packet overhead.
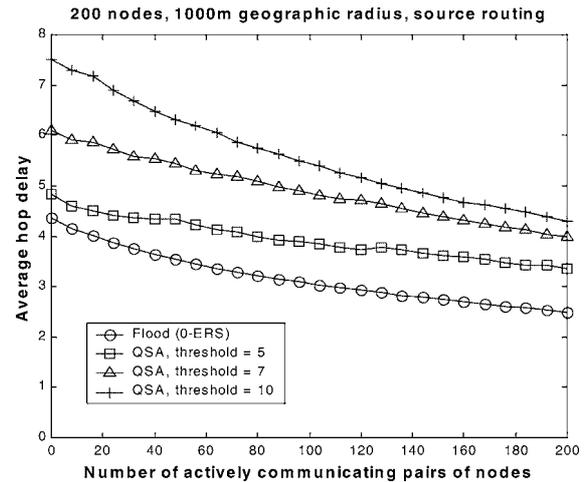
Figure 16. 200 nodes, hop delay.

obsolete by the time it reaches s. The snapshot model, therefore, relaxes the simulator requirements, as mobility effects need not be considered for the purposes of assessing route discovery overhead.

Three other simulation assumptions concern the communication sessions between pairs of nodes. One, mentioned already in section 2.1, is that each packet contains the path length, of the route used between s and t, in its datagram header. Another is that each communication session consists of unicast communication(s) between a unique (unordered) pair of nodes whose datagrams are forwarded over a single, least-hop, bi-directional network path (i.e., alternate paths not used). Lastly, for each communication session, packets are originated at a steady rate so that neighbors of nodes lying on an active path are able to use the overheard packet transmission to refresh their route caches.

Avoiding large inter-packet times can be facilitated without overhead if the communication session is for a CBR application. In the case of a bursty communication session, s can originate "heartbeat" packets to notify tapping nodes along an active path that the topology data they have cached about the route is still fresh. Implied in the consideration here of the communication sessions, is the notion of avoiding stale route cache entries. That is, the nature of the communication sessions ensures that topology data obtained via packet tapping is valid, provided it is *purged* from the route cache when the communication session, from which it was learned, is either terminated or resorts to a different path. Purging route cache entries when the associated communication session is considered no longer active is consistent with the *Link-Static-x* cache purging approach analyzed in [4], where x is the expiration timeout period.

4.2. Discussion of results

The results reported in figures 11–16 demonstrate the responsiveness of the QSA to user-specified maximum allowable hop delay. That is, when η_{max} is small, the QSA employs a coarse-grained form of ERS to keep η low. However, as η_{max}

is increased, the QSA selects a gradual form of ERS with increasing likelihood so as to reduce ψ. This has been achieved for node counts ranging from 50 to 200 nodes, which represents a significant dynamic range of successful operation.

It is worth emphasizing that the curves of figures 11–16 report *average* packet overhead and hop delay. This is of particular relevance in terms of satisfying the $\eta_{\max}$ hop delay bound. That is, although the QSA successfully maintains a hop delay that is less than $\eta_{\max}$, on average, there may be instances where a particular target node will incur route acquisition delay that is considerably larger than that predicted by the QSA and may even exceed $\eta_{\max}$. Further, there may exist instances of network configurations where estimations of P, L and $|N_k|$ are inherently inaccurate, and may result in the QSA selecting an ERS approach that incurs an average hop delay in excess of $\eta_{\max}$. Nevertheless, demonstration that the mean route acquisition delay can be bounded by the QSA represents an important result.

Not shown explicitly in figures 11–16 is the phenomenon of the QSA transitioning to a gradual form of ERS with increasing P. The decision to employ gradual ERS with increasing likelihood as P increases is also compatible with prevailing traffic conditions. For example, when the path count is large (e.g., $|V|$ paths), then with high probability routing information will be cached locally near s and gradual ERS makes sense. This tends to reduce ψ which is crucial for large P as the network is already experiencing significant traffic load due to communication sessions between users assuming, of course, user traffic volume is proportional to the number of communication sessions. On the other hand, when the path count is small (e.g., 0 paths) then routing information will be cached locally near s with small probability and immediate network-wide flood search makes sense. This may incur up to $|V| - 1$ RREQ packet transmissions. However, since the network is lightly loaded by user traffic, a single network-wide flooding event is not likely to precipitate congestion.

5. Contributions and conclusions

The primary contributions of this paper are as follows.

- A novel architecture, with low communication and computation overhead, has been proposed for dynamically selecting the ERS. This architecture is applicable to both source routing (e.g., DSR) and next-hop routing (e.g., AODV) protocols.

- A *robust analytical model* that characterizes nodes in N_k and X as sub-regions of the network area has been formulated.

- New techniques for estimating $\mathrm{E}[H]$, $|N_k|$, $p_{j,k}$ and η have been devised.

- Simulations demonstrate proof of concept.

Of the above contributions, worthy of additional discussion is the motivation behind the robust analytical model. This model for 2-dimensional MANETs permits the incremental probability of successful route acquisition, $p_{j,k}$, to

be calculated in terms of the probability that the geographical area covered by nodes in X overlaps the area covered by the nodes in $N_{k+1} - N_{j+1}$, as shown in figure 10. Because it is consistent with the configuration that actually occurs in *2-dimensional networks*, it allows the QSA proposed here to reliably select an ERS approach that reduces ψ while satisfying the $\eta_{\max}$ bound, on average. Further, because of its analytical nature, the model supports *efficient* real time computation of $p_{j,k}$ via (6)–(8). Thus, it satisfies the potentially conflicting requirements of accurately representing actual network conditions while providing a means to efficiently to compute $p_{j,k}$. Lastly, due to the considerable dependence of $p_{j,k}$ on an accurate estimate of $|N_k|$, a detailed procedure to account for the effect of the network geographic boundary on k-hop neighborhood node counts is proposed in the Appendix. Here again, this portion of the analytical model reflects the network environment that would occur in practice.

The assumption that $\tau \propto \eta$ is possibly a conservative simplification in that τ may be super-linear in η. This, however, means that the results reported herein actually represent a conservative depiction of the benefits of employing a QSA. That is, if $\tau \propto \eta^{1+\alpha}$ ($\alpha > 0$) rather than $\tau \propto \eta$, then it is even more important to employ a QSA since the variance of τ, depending on the type of ERS chosen, will be greater than indicated here. More accurate modeling of the relationship between τ and η represents a direction for future work.

Other future work on the QSA proposed here includes consideration of additional ERS options (e.g., *1-4-ERS*). Also, the efficacy of the QSA in satisfying the $\eta_{\max}$ requirement and its accuracy in estimating η remain to be assessed. Another aspect of the QSA to study further is its performance *without* refined estimates of P and L. Such an implementation would relax the requirements on the Hello protocol. Lastly, a QSA extension to account for a non-circular network geometry should be investigated.

Appendix

This Appendix presents a more robust procedure for estimating the k-hop neighborhood size. In the overview presented in section 3.2, it is assumed that a node v is situated at the geographic center of the network area. Here, however, it is assumed that v is situated randomly somewhere in a circular area of radius R'_{G}, where R'_{G} is an estimate of the actual geographic radius of the network R_{G}. R'_{G} is computed as follows:

$$R'_{\mathrm{G}} = \sqrt{\frac{|V| \cdot R_{\mathrm{TX}}^2}{1 + \deg(v)}}. \qquad \text{(A.1)}$$

Given the assumption of a uniform 2-dimensional random distribution on node location, the average distance ($\mathrm{E}[R]$) from the geographic center of the network is given simply as follows:

$$\mathrm{E}[R] = R'_{\mathrm{G}} \int_0^1 \sqrt{y}\, \mathrm{d}y = \frac{2}{3} R'_{\mathrm{G}}. \qquad \text{(A.2)}$$

This condition is illustrated in figure 17.
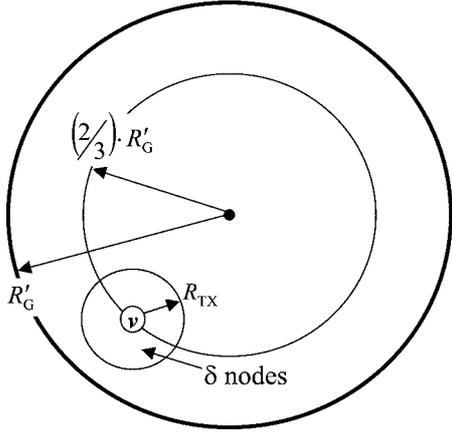
Figure 17. Intersection of area of coverage with actual network area when $r_k \leqslant 1/3$.

Figure 18. Intersection of area of coverage with actual network area when $1/3 < r_k < \sqrt{5}/3$.

Next, it is recalled that R_k is defined as the estimate for the geographic radius of the k-hop neighborhood about v, as considered in (3). Now, r_k is defined as the ratio of R_k to R'_G:

$$r_k = \frac{R_k}{R'_G}. \qquad (A.3)$$

As illustrated in figures 17–21, there are a total of five distinct scenarios that may arise when considering the area of intersection between the geographic coverage of N_k and the actual network area. In each of these figures, the estimated network radius (R'_G) has been normalized to 1. These cases are as follows:

- Case 1: $r_k \leqslant 1/3$ (figure 17);
- Case 2: $1/3 < r_k < \sqrt{5}/3$ (figure 18);
- Case 3: $\sqrt{5}/3 \leqslant r_k \leqslant \sqrt{13}/3$ (figure 19);
- Case 4: $\sqrt{13}/3 < r_k < 5/3$ (figure 20);
- Case 5: $r_k \geqslant 5/3$ (figure 21).

The range of values between $\sqrt{5}/3$ and $\sqrt{13}/3$ for case 3 correspond to values of r_k for which $A_2 = 0$. Referring to figure 18, $A_2 > 0$ for $\theta < \pi/4$ radians. As shown in figure 22, however, it is evident that $A_2 = 0$ at $\theta = \pi/4$ radians.

Figure 19. Intersection of area of coverage with actual network area when $\sqrt{5}/3 \leqslant r_k \leqslant \sqrt{13}/3$.

Figure 20. Intersection of area of coverage with actual network area when $\sqrt{13}/3 < r_k < 5/3$.
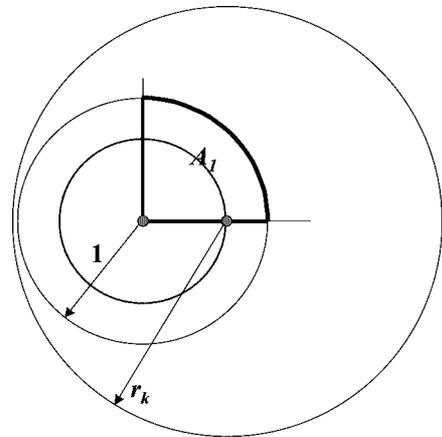
Figure 21. Intersection of area of coverage with actual network area when $r_k \geqslant 5/3$.
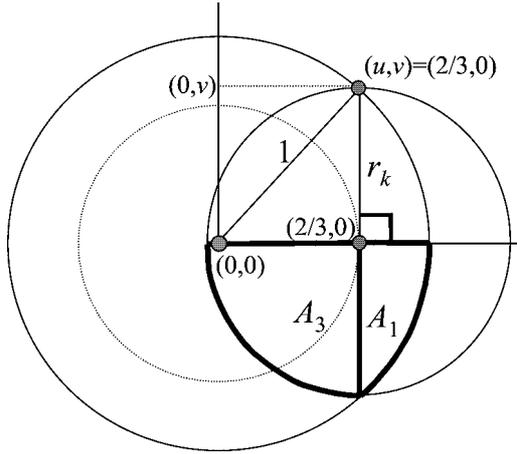
Figure 22. Deriving the limit $\sqrt{5}/3$ for cases 2 and 3.
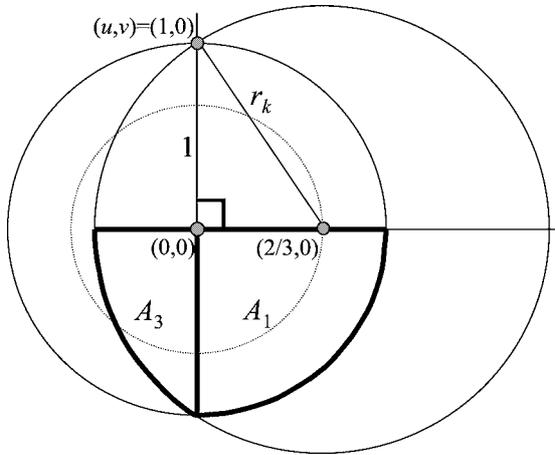
Figure 23. Deriving the limit $\sqrt{13}/3$ for cases 3 and 4.

By applying the Pythagorean Theorem and recalling that the circle of k-hop query coverage is presumed centered at a location that is $2R_G/3$ units from the center of the network: $(2/3)^2 + r_k^2 = 1^2 \Rightarrow r_k = \sqrt{5}/3$. The limit of $\sqrt{13}/3$ is computed in a similar fashion, as shown in figure 23. Here, the right angle is formed at the center of the network area where the legs of the right triangle are known ($2/3$ and 1) and it is the hypotenuse that is computed via the Pythagorean Theorem: $(2/3)^2 + 1^2 = r_k^2 \Rightarrow r_k = \sqrt{13}/3$.

For cases 1 and 5, the area of intersection is trivial to compute (i.e., $\pi \cdot r_k^2$ and π, respectively). For cases 2–4, although more complex, the area of intersection may be computed via integral calculus. As an example, the area of intersection for case 2 is computed as follows. Referring to figure 18, the area intersection (A) is clearly given by:

$$A = 2(A_1 + A_2 + A_3). \tag{A.4}$$

A_3 is computed trivially ($\pi \cdot r_k^2/4$). The computation of A_1 and A_2 are based on the geometry illustrated in figures 24 and 25, respectively:

$$A_1 = \int_{d_1/2}^{1} \sqrt{1 - x^2}\, dx \tag{A.5a}$$
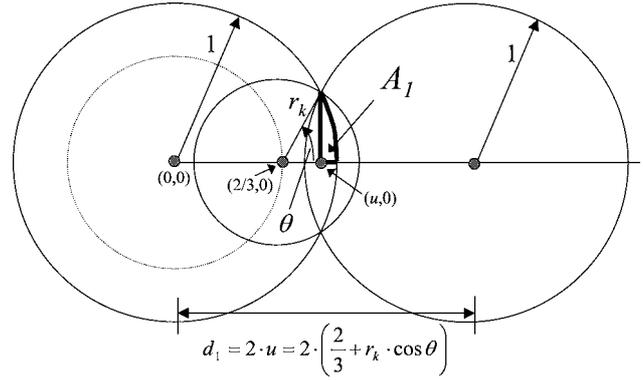
Figure 24. Computation of A_1.
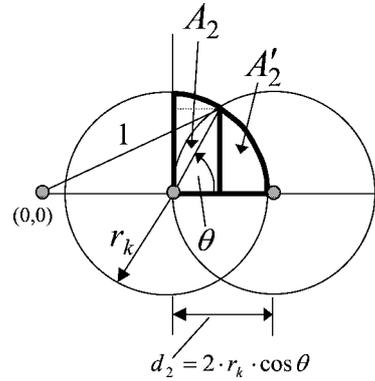
Figure 25. Computation of A_2.
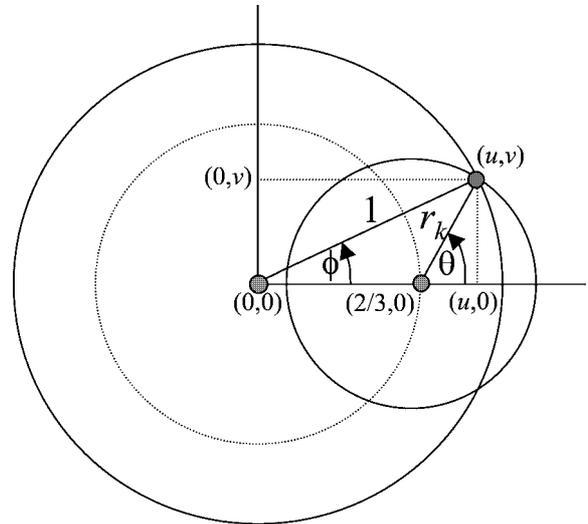
Figure 26. Geometry for computing θ.

$$A_2 = \frac{\pi \cdot r_k^2}{4} - \int_{d_2/2}^{r_k} \sqrt{r_k^2 - x^2}\, dx. \tag{A.5b}$$

The distances d_1 and d_2 can be obtained via application of trigonometry. Figure 26 depicts the geometric representation of the simulated environment. Again, R_G' has been normalized to unity. Applying trigonometry to figure 26 yields the following relations:

$$u = \cos\phi = \frac{2}{3} + r_k \cdot \cos\theta, \tag{A.6a}$$

$$v = \sin\phi = r_k \cdot \cos\theta, \tag{A.6b}$$

$$u^2 + v^2 = 1 \tag{A.7a}$$

$$\Rightarrow \left(\frac{2}{3} + r_k \cdot \cos\theta\right)^2 + (r_k \cdot \sin\theta)^2 = 1 \tag{A.7b}$$

$$\Rightarrow r_k^2 + \frac{4r_k}{3}\cos\theta + \frac{4}{9} = 1 \tag{A.7c}$$

$$\Rightarrow \cos\theta = \frac{3}{4r_k}\left(\frac{5}{9} - r_k^2\right) \tag{A.7d}$$

$$\Rightarrow \theta = \cos^{-1}\left[\frac{3}{4r_k}\left(\frac{5}{9} - r_k^2\right)\right]. \tag{A.7e}$$

The distances d_1 and d_2 employed as limits of integration in (A.5a) and (A.5b), respectively, are then computed for each value of r_k:

$$d_1 = 2 \cdot \left(\frac{2}{3} + r_k \cdot \cos\theta\right), \tag{A.8a}$$

$$d_2 = 2r_k \cdot \cos\theta. \tag{A.8b}$$

Lastly, planar geometry, trigonometry and integral calculus are applied to compute the area of intersection for cases 3 and 4. For all cases, because the radius of the circular areas in question are normalized with respect to the estimated geographic radius (R'_G), A lies in the range $0 < A \leqslant \pi$ (for cases 1–4). Once A is computed via (A.4), the estimates of $|N_k|$ computed via (3) are revised as follows:

$$|N_k|'' = \frac{|N_k|' \cdot A}{\pi}. \tag{A.9}$$

The revised estimates of $|N_k|$ given by (A.9) are used in the estimations of $E[H \mid H \geqslant 1, \ P = 0]$ in (4), $f_{X|j}$ in (6) and $f_{j,k}$ in (7).

Acknowledgements

References

[1] J. Broch, D.A. Maltz, D.A. Johnson, Y.-C. Hu and J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, in: *Proc. of ACM/IEEE MobiCom* (October 1998) pp. 85–97.

[2] R. Castaneda and S.R. Das, Query localization techniques for on-demand routing protocols in ad hoc networks, in: *Proc. of ACM/IEEE MobiCom* (August 1999) pp. 186–194.

[3] S.R. Das, C.E. Perkins and E.M. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks, in: *Proc. of IEEE Infocom* (April 2000) pp. 3–12.

[4] Y.-C. Hu and D.B. Johnson, Caching strategies in on-demand routing protocols for wireless networks, in: *Proc. of ACM/IEEE MobiCom* (August 2000) pp. 231–242.

[5] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla and T.-W. Chen, Scalable routing strategies for ad hoc wireless networks, IEEE J. Select. Areas Commun. 17(8) (1999) 1369–1379.

[6] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek and M. Degermark, Scenario-based performance analysis of routing protocols for mobile ad-hoc networks, in: *Proc. of ACM/IEEE MobiCom* (August 1999) pp. 195–206.

[7] D.B. Johnson and D.A. Maltz, Dynamic source routing in ad hoc wireless networks, in: *Mobile Computing*, eds. T. Imielinski and H. Korth (Kluwer Academic, 1996) ch. 5, pp. 153–181.

[8] D.B. Johnson, D.A. Maltz, Y.-C. Hu and J.G. Jetcheva, The dynamic source routing protocol for mobile ad hoc networks, IETF Internet Draft, draft-ietf-manet-dsr-07.txt (21 February 2002).

[9] B. Karp and H.T. Kung, GPSR: Greedy perimeter stateless routing for wireless networks, in: *Proc. of ACM/IEEE MobiCom* (August 2000) pp. 243–254.

[10] Y.-B. Ko and N. Vaidya, Location-aided routing (LAR) in mobile ad hoc networks, in: *Proc. of ACM/IEEE MobiCom* (October 1998) pp. 66–75.

[11] J. Li, J. Janotti, D. De Couto, D. Karger and R. Morris, A scalable location service for geographic ad-hoc routing, in: *Proc. of ACM/IEEE MobiCom* (August 2000) pp. 120–130.

[12] D.A. Maltz, J. Broch, J. Jetcheva and D.B. Johnson, The effects of on-demand behavior in routing protocols for multihop wireless ad hoc networks, IEEE J. Select. Areas Commun. 17(8) (1999) 1439–1453.

[13] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen and J.-P. Sheu, The broadcast storm problem in a mobile ad hoc network, in: *Proc. of ACM/IEEE MobiCom* (August 1999) pp. 151–162.

[14] W. Peng and X.-C. Lu, On the reduction of broadcast redundancy in mobile ad hoc networks, in: *Proc. of MobiHoc Workshop* (August 2000) pp. 129–130.

[15] C.E. Perkins and E. M. Royer, Ad-hoc on-demand distance vector routing, in: *Proc. of IEEE WMCSA '99* (February 1999).

[16] C.E. Perkins, E.M. Royer and S.R. Das, Ad hoc on-demand distance vector (AODV) routing, IETF Internet Draft, draft-ietf-manet-aodv-11.txt (19 June 2002).

[17] J. Sucec and I. Marsic, An application of parameter estimation to route discovery by on-demand routing protocols, in: *Proc. of IEEE ICDCS '01* (April 2001) pp. 207–216.

[18] Y.-C. Tseng, S.-Y. Ni and E.-Y. Shih, Adaptive approaches to relieving broadcast storms in a wireless multihop mobile ad hoc network, in: *Proc. of IEEE ICDCS '01* (16–19 April 2001) pp. 481–488.

John Sucec is a Ph.D. student in the Department of Electrical and Computer Engineering at Rutgers University. Previously, he studied at the University of Connecticut and Binghamton University. His research interests include communication protocols for datagram forwarding networks.
E-mail: jsucec@caip.rutgers.edu

Ivan Marsic is an Assistant Professor of Electrical and Computer Engineering at Rutgers University. He received the B.S. and M.S. degrees in computer engineering from University of Zagreb, Croatia, in 1982 and 1987, respectively, and Ph.D. degree in biomedical engineering from Rutgers University, New Brunswick, NJ, in 1994. Prof. Marsic's research interests are in mobile computing, groupware, computer networks, and multimodal human–computer interfaces. He has authored over 80 journal and conference papers, one book and 3 book chapters. Prof. Marsic is a member of IEEE, ACM, IFIP WG2.7/13.4, and International Academy for Advanced Decision Support and has been a consultant to industry and government.

E-mail: marsic@caip.rutgers.edu