The Performance of Query Control Schemes for the Zone Routing Protocol♥

Zygmunt J. Haas
School of Electrical Engineering, Cornell University
Ithaca, NY 14853
Tel: (607) 255-3454
haas@ee.cornell.edu

Marc R. Pearlman
School of Electrical Engineering, Cornell University
Ithaca, NY 14853
Tel: (607) 254-8816
pearlman@ee.cornell.edu

ABSTRACT

In this paper, we study the performance of route query control mechanisms for the recently proposed Zone Routing Protocol (ZRP) for ad-hoc networks. The ZRP proactively maintains routing information for a local neighborhood (routing zone), while reactively acquiring routes to destinations beyond the routing zone. This hybrid routing approach has the potential to be more efficient in the generation of control traffic than traditional routing schemes. However, without proper query control techniques, the ZRP can actually produce more traffic than standard flooding protocols.

Our proposed query control schemes exploit the structure of the routing zone to provide enhanced detection (Query Detection (QD1/QD2)), termination (Loop-back Termination (LT), Early Termination (ET)) and prevention (Selective Bordercasting (SBC)) of overlapping queries. We demonstrate how certain combinations of these techniques can be applied to single channel or multiple channel ad-hoc networks to improve both the delay and control traffic performance of the ZRP. Our query control mechanisms allow the ZRP to provide routes to all accessible network nodes with only a fraction of the control traffic generated by purely proactive distance vector and purely reactive flooding schemes, and with a response time as low as 10% of a flooding route query delay.

Keywords

Ad-hoc networks, routing, protocols, Zone Routing, routing zone

1. INTRODUCTION

An ad-hoc network is a self-organizing wireless network made up of mobile nodes and requiring no fixed infrastructure. The limitations on power consumption imposed by portable wireless radios result in a node transmission range that is typically small relative to the span of the network. To provide communication throughout the entire network, nodes are designed to serve as relays if needed. The result is a distributed multi-hop network with a time-varying topology.

Because ad-hoc networks do not rely on existing infrastructure and are self-organizing, they can be rapidly deployed to provide robust communication in a variety of hostile environments. This makes ad-hoc networks very appropriate for providing tactical communication for military, law enforcement and emergency response efforts. Ad-hoc networks can also play a role in civilian forums such as the electronic classroom, convention centers and construction sites. With such a broad scope of applications, it is not difficult to envision ad-hoc networks operating over a wide range of coverage areas, node densities and node velocities.

This potentially wide range of ad-hoc network operating configurations poses a challenge for developing efficient routing protocols. On one hand, the effectiveness of a routing protocol increases as network topology information becomes more detailed and up-to-date. On the other hand, in an ad-hoc network, the topology may change quite often, requiring large and frequent exchanges of data among the network nodes. This is in contradiction with the fact that all updates in the wireless communication environment travel over the air and are costly in resources.

Existing routing protocols can be classified either as *proactive* or as *reactive.* Proactive protocols attempt to continuously evaluate the routes within the network, so that when a packet needs to be forwarded, the route is already known and can be immediately used. The family of Distance-Vector protocols is an example of a proactive scheme. Early applications of proactive routing schemes for ad-hoc networks were based on the Distributed Bellman-Ford (DBF) algorithm [1]. Modifications to the basic DBF algorithm (i.e. [2], [3] and [11]) were proposed to address issues such as convergence and excessive traffic. More recently, new proactive protocols have been designed for packet radio networks, such as the Wireless Routing Protocol (WRP) [8] and [9].

Reactive protocols, on the other hand, invoke a route determination procedure on an as-needed basis. Typically, reactive routing protocols are based on query-reply exchanges. The classical flooding algorithms are examples of reactive protocols. Reactive protocols have also been specifically designed for the ad-hoc network environment, such as [6], [10] and [12].

The advantage of the proactive schemes is that route information is available when needed, resulting in little delay prior to data transmission. In contrast, reactive schemes may produce

♥ This work is supported by the US Air Force/Rome Labs, under the contract number C-7-2544 and a grant from Motorola Corporation, the Applied Research Laboratory

significant delay in order to determine a route when route information is needed, but not available.

Routing schemes, whether proactive or reactive, require some exchange of control traffic. This overhead can be quite large in ad-hoc networks, where the topology frequently changes. Reactive protocols produce a large amount of traffic by effectively flooding the entire network with route queries. The combination of excessive control traffic and long route query response time rule out pure reactive routing protocols for real-time communication applications. Pure proactive schemes are likewise not appropriate for ad-hoc networks, as they *continuously* use a large portion of the network capacity to keep the routing information current. Proactive protocols tend to distribute topological changes widely in the network, even though the creation/destruction of a new link at one end of the network may not be a significant piece of information at the other end of the network. Furthermore, since ad-hoc network nodes may move quite fast, and as the changes may be more frequent than the route requests, most of this maintained routing information is never used! This results in further waste of the network capacity.

2. THE ZONE ROUTING PROTOCOL – A SHORT OVERVIEW

The behavior of purely proactive and reactive schemes suggest that what is needed is a protocol that initiates the route-determination procedure on-demand, but at limited search cost. Our protocol, the *Zone Routing Protocol (ZRP)* ([4] and [5]), is an example of such a hybrid reactive/proactive scheme. On one hand, it limits the scope of the proactive procedure only to the node's local neighborhood. As we shall see, the local routing information is referred to quite often in the operation of the ZRP, minimizing the waste associated with the purely proactive schemes. On the other hand, the search throughout the network, although it is global, can be performed efficiently by querying selected nodes in the network, as opposed to querying all the network nodes.

We proceed with an introduction of the routing zone concept and a brief overview of the ZRP architecture.

2.1 The Notion of a Routing Zone and Intrazone Routing

In the ZRP, a node proactively maintains routes to destinations within a local neighborhood, which we refer to as a *routing zone*. More precisely, a node's routing zone is defined as a collection of nodes whose minimum distance in hops from the node in question is no greater than a parameter referred to as the *zone radius*. Note that each node maintains its own routing zone. An important consequence, as we shall see, is that the routing zones of neighboring nodes overlap.

Figure 1 illustrates the routing zone concept with a routing zone of radius 2 hops. This particular routing zone belongs to node S, which we refer to as the *central node* of the routing zone. Nodes A through K are members of S's routing zone. Node *L*, however, is three hops away from S, and is therefore outside of *S*'s routing zone. An important subset of the routing zone nodes is the collection of nodes whose minimum distance to the central node is exactly equal to the zone radius. These nodes are aptly named peripheral nodes. In our example, nodes G-K are peripheral nodes of node S. We typically illustrate a routing zone as a circle centered around the central node. However, one should keep in mind that the zone is not a description of physical distance, but rather nodal connectivity (hops).

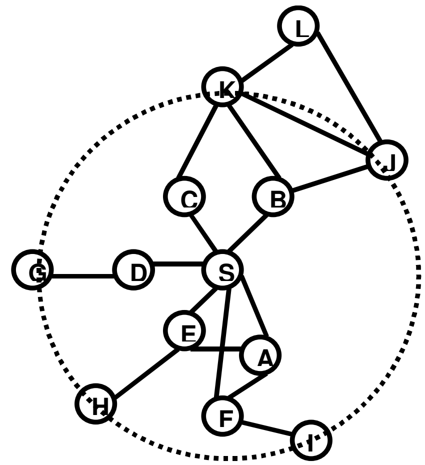

Figure 1: A Routing Zone of Radius 2 Hops

The construction of a routing zone requires a node to first know who its *neighbors* are. A neighbor is defined as a node that can communicate directly with the node in question[1] (and is thus one hop away). Identification of a node's neighbors may be provided directly by the media access control (MAC) protocols, as in the case of polling-based protocols. In other cases, neighbor discovery may be implemented through a MAC-level *Neighbor Discovery Protocol (NDP).* Such a protocol typically operates through the periodic broadcasting of "hello" beacons. The reception (or quality of reception) of a "hello" beacon can be used to indicate the status of a connection to the beaconing neighbor.

The ZRP maintains routing zones through a proactive component called the IntrAzone Routing Protocol (IARP). In this paper, the IARP is implemented as a modified distance vector scheme. However, almost any distributed proactive routing scheme may be modified to perform the role of the IARP. Our experience suggests that the tradeoffs are not strongly affected by the particular choice of the proactive scheme used. Changes in neighbor connectivity trigger the broadcast of IARP route update packets reflecting the new status to the neighbor. In the case of a neighbor discovery, the node also transmits a copy of its routing table to the new neighbor. A node that receives an IARP route update packet will record the updated information and determine if the update results in a change to the shortest path to the destination. If so, the node will broadcast an IARP route update reflecting this change. Whereas regular proactive schemes will allow updates to propagate throughout the entire network, the IARP uses the route hop count to prevent updates from propagating farther than the zone radius.

[1] The determination of a direct connection between two nodes is typically based on measurements of link quality, such as received signal power, bit error rate (BER), signal to interference ratio (SIR), link stability, etc. The application of the network often determines the minimal level of link quality to support a direct connection between two nodes.

2.2 Interzone Routing

Whereas the IARP maintains routes to nodes within the routing zone, the *IntErzone Routing Protocol (IERP)* is responsible for acquiring routes to destinations that are located beyond the routing zone. The IERP uses a query-response mechanism to discover routes on demand.

The IERP is distinguished from standard flooding algorithms by exploiting the structure of the routing zone, through a process known as *bordercasting*. Bordercasting is a packet delivery service that allows a node to send a message to its peripheral nodes. The ZRP provides this service through a component called the Bordercast Resolution Protocol (BRP).

In its simplest form, bordercasting may be implemented through network layer unicasting or multicasting of messages to the peripheral nodes. This approach prevents non-peripheral nodes from accessing the bordercasted messages as they are relayed to the edge of the routing zone. As will be shown later, such access is central to the control of the route query process. As such, a more suitable implementation of bordercasting indirectly sends messages to peripheral nodes by forwarding between BRP layers of adjacent nodes. In such an implementation, either IP unicast or IP broadcast is used as the network layer delivery mechanism.

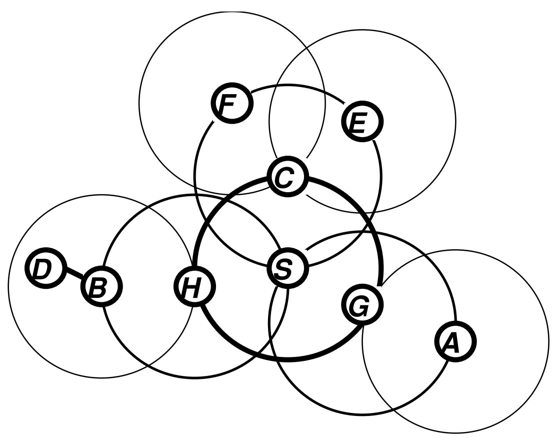

Figure 2: An Example of IERP Operation

An IERP route query is triggered by the network layer, when a data packet is to be sent to a destination that does not lie within its routing zone[2]. The source generates a route query packet, which is uniquely identified by a combination of the source node's ID and request number. The query is then bordercast to all the source's peripheral nodes. Upon receipt of a route query packet, a node adds its ID to the query. The sequence of recorded node IDs specifies an *accumulated route* from the source to the current routing zone.[3] If the destination does not appear in the node's routing zone, the node bordercasts the query to *its* peripheral nodes[4]. If the destination is a member of the routing zone, a *route reply* is sent back to the source, along the path specified by reversing the accumulated route. As with standard flooding algorithms, a node will discard any route query packet for a query that it has previously encountered.

An example of this *Route Discovery* procedure is demonstrated in Figure 2. The source node S prepares to send a data to the destination D. S first checks whether D is within its routing zone. If so, S already knows the route to node *D.* Otherwise, *S* sends a query to all its peripheral nodes (*C, G,* and *H).* Now, in turn, each one of these nodes, after verifying that *D* is not in its routing zone forwards the query to its peripheral nodes. In particular, *H* sends the query to *B,* which recognizes *D* as being in its routing zone and responds to the query, indicating the forwarding path: *S-H-B-D.*

A nice feature of this route discovery process is that a single route query can return multiple route replies. The quality of these returned routes can be determined based on hop count (or any other path metric[5] accumulated during the propagation of the query). The best route can be selected based on the relative quality of the route (i.e. choose the route with the smallest hop count, or shortest accumulated delay).

The inter-relationship of the ZRP component protocols is illustrated in Figure 3. The proactive maintenance of the routing zone topology is performed by the IARP, through exchange of route update packets. Route updates are triggered by the MAC-level NDP, which notifies the IARP when a link to a neighbor is established or broken. The IERP reactively acquires routes to nodes beyond the routing zone using a query-reply mechanism. The IERP forwards queries to its peripheral nodes through a bordercast delivery service provided by the BRP. The BRP keeps track of the peripheral nodes through up-to-date routing zone topology information provided by the IARP. The IERP also makes use of the IARP routing zone information to determine whether a queried for destination belongs to its routing zone.

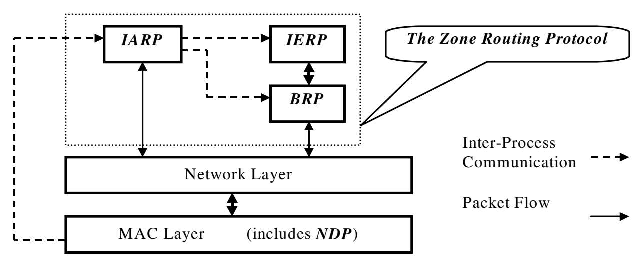

Figure 3: ZRP's Architecture

The relationship between the IARP and IERP may, at first, give the impression that the ZRP is a hierarchical routing protocol. In fact, the ZRP bears only a superficial resemblance to such protocols. Hierarchical routing relies on the strategic assignment of gateways or landmarks[13] to establish a hierarchy of subnets for the entire network.[6] Access to a subnet is provided through

---

[2] Remember that a node knows the identity, distance to, and a route to all the nodes in its zone.

[3] Because each node maintains a routing zone, interzone routes can be specified as a sequence of nodes separated by a distance equal to the zone radius.

[4] Excluding the peripheral node from which the route query was received.

[5] Typical path metrics include hop count, delay, capacity, etc.

[6] Gateways or landmarks must be assigned in such a way that every node is able to access every level of the hierarchy. Furthermore, in order to guarantee communication between any two network nodes, there must be a "top" subnet or landmark which is accessible / visible by all network nodes.

that subnet's assigned gateway or landmark. As a result, two nodes that belong to different subnets must send their communication up the hierarchy to a subnet which is common to both nodes. This constraint often leads to sub-optimal routes. In contrast, access to a ZRP routing zone is provided not through a single gateway or landmark, but through the "best" of the multiple peripheral nodes that define the extent of the zone. Communication beyond a routing zone is passed *across* overlapping routing zones in a peer-to-peer manner, rather than up to a higher tier with broader coverage. As a result, the routing inefficiencies associated with hierarchical routing protocols are avoided in the ZRP, permitting optimal routing to a destination. In this sense, it may actually be more accurate to categorize the ZRP as a flat rather than a hierarchical routing protocol.

3. QUERY CONTROL MECHANISMS

Because the routing zones heavily overlap, the route query will be forwarded to many network nodes, multiple times. In fact, it is very possible that the query will be forwarded to all the network nodes, effectively flooding the network. But a more disappointing result is that, due to fact that bordercasting involves sending the query over a path of length equal to the zone radius, the IERP will result in much more traffic than the flooding itself! What is needed is a more efficient termination criterion than the standard flooding algorithms provide.

In order to understand the cause of the ZRP control traffic problem, it is important to stress one of the key features of the routing zone: A node's response to a route query contains information about that node's entire routing zone. From this perspective, excess route query traffic can be regarded as a result of overlapping query threads (i.e. overlapping queried routing zones). Thus, the design objective of query control mechanisms should be to reduce the amount of route query traffic by steering threads outward from the source's routing zone and away from each other (see Figure 4). This problem is addressed from two different perspectives: thread overlap detection/termination and thread overlap prevention.

The standard approach to query thread termination is to discard a thread when it appears at a previously queried node. However, this does not fully exploit the structure of the routing zone. A broader approach is to discard a thread that appears in a previously queried zone. This criterion introduces the first challenge for the design of an effective termination mechanism: How can a previously queried zone be identified when only one node (the central node) was queried?

3.1 Loop-back Termination (LT)

It is relatively easy to identify a thread that returns to a routing zone that it previously queried. A node simply examines the accumulated route in the received route query packet to determine if any hop (excluding the most recent hop) lies within its routing zone. If the loop-back condition exists, the thread is discarded. An example of this scheme, which we refer to as Loop-back Termination (LT), is shown in figure 5. Node S bordercasts a route query to A, which bordercasts it to B, which in turn bordercasts it to C. C terminates the thread (i.e. does not bordercast) because node S, which appears in the accumulated route, also lies within C's routing zone. LT is an ideal mechanism to handle thread loop-back, because the information provided by the accumulated route is sufficient to identify all cases of loop-back.

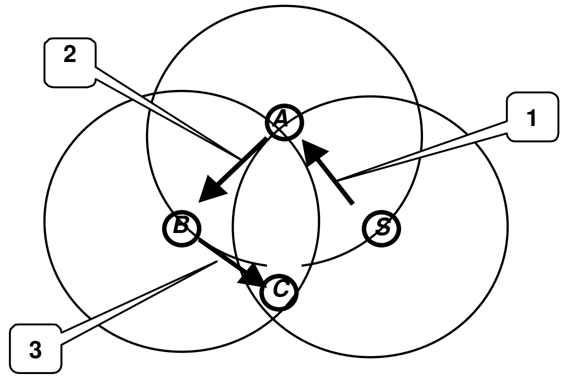

Figure 5: Loop-back Termination (LT)

3.2 Query Detection (QD1/QD2)

A majority of thread overlapping occurs by a thread appearing in a zone that was previously queried by another thread. Unlike the loop-back case just described, the ability to terminate in this situation strongly depends on the ability of nodes to detect that a routing zone which they belong to has been previously queried[7]. Clearly, the central node in the routing zone (which processed the query) is aware that its zone has been queried. In order to notify the remaining routing zone nodes, without introducing additional control traffic, some form of "eavesdropping" needs to be implemented. Based on the ZRP architecture described earlier, it is most convenient to perform query detection at the BRP, which is responsible for query delivery. The first level of Query Detection (QD1), allows the intermediate nodes, which transport queries to the edge of the routing zone, to detect these queries. In single channel networks, it may be possible for queries to be detected by any node within the range of a query-transmitting node. This extended query detection capability (QD2) can be implemented by using IP broadcasts to send route queries[8]. Figure 6 illustrates both levels of advanced query detection. In this

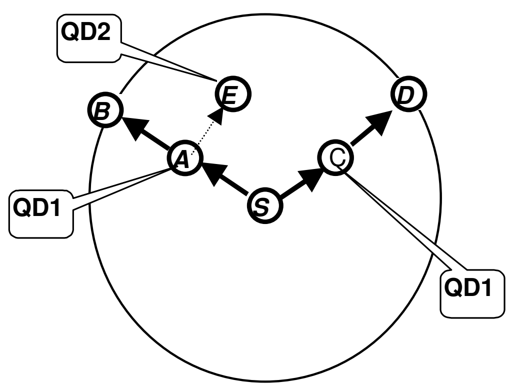

Figure 6: Advanced Query Detection (QD1/QD2)

[7] The ID of the node that bordercasted the first detected query thread is also recorded. In order to ensure full network coverage for that query, future threads received from that bordercasting node are <u>not</u> automatically discarded.

example, node S bordercasts to two peripheral nodes, B and D. The intermediate nodes A and C are able to detect passing threads using QD1. If QD2 is implemented, node E will be able to "eavesdrop" on A's transmissions and record the query as well.

3.3 Early Termination (ET)

The termination criteria can be further tightened by discarding a thread as it enters a previously queried zone. When the ability to terminate threads is limited to peripheral nodes, threads are allowed to penetrate *into* previously covered areas, generating unnecessary control traffic. This excess traffic can be eliminated by extending the thread termination capability to the intermediate nodes that transport the thread. We refer to this approach as Early Termination (ET). Figure 7 illustrates the operation of the ET mechanism. Node S bordercasts a route query, with node C as one of the intended recipients. Intermediate node A passes along the query to B. Instead of delivering the query to node C, node B terminates the thread because a different thread of this query was previously detected. It should be noted that ET only allows partial participation of intermediate nodes in the route query process. Intermediate nodes are restricted from issuing new queries. Otherwise, the ZRP would degenerate into a flooding protocol.

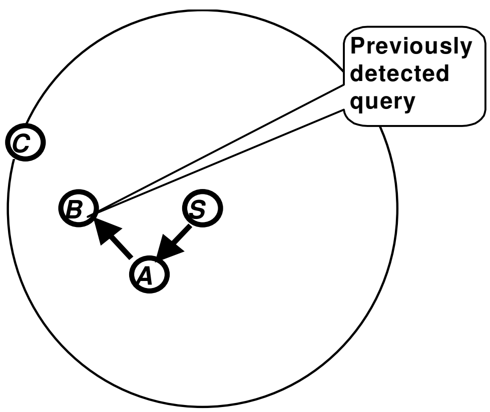

Figure 7: Early Termination (ET)

4. Selective Bordercasting (SBC)

We now address the more complicated issue of thread overlap prevention. By concentrating on the elimination of overlap locally, some degree of control can be imposed on the direction of thread propagation, thereby reducing thread overlap farther out in the network. Local thread overlap is due to the heavy overlap of peripheral nodes' routing zones, especially as the routing zone radius increases. Rather than bordercast queries to all peripheral nodes, the same coverage can be provided by bordercasting to a properly chosen subset of peripheral nodes, through a mechanism that we term Selective Bordercasting (SBC).

SBC requires that the IARP provides network topology information for an extended zone that is twice the radius of the routing zone. Prior to bordercasting, a node first determines the subset of outer peripheral nodes[9] covered by its assigned inner peripheral nodes. The node then bordercasts to a subset of the assigned inner peripheral nodes which form a "minimal" partitioning set of the outer peripheral nodes. Figure 8 provides an illustrative example of a SBC application. Node S's inner peripheral nodes are A, B and C. Node S's outer peripheral nodes are F, G, H, X, Y and Z. We can see from the overlapping routing zones that the two inner peripheral nodes of node B (H and X) are also inner peripheral nodes of A and C. Consequently, node S can choose to eliminate node B from its list of bordercast recipients. Node A can provide coverage to F, G and H, and node C can cover X,Y and Z. In this way, we are able to provide full coverage over the extended zone while preventing overlapping queries.

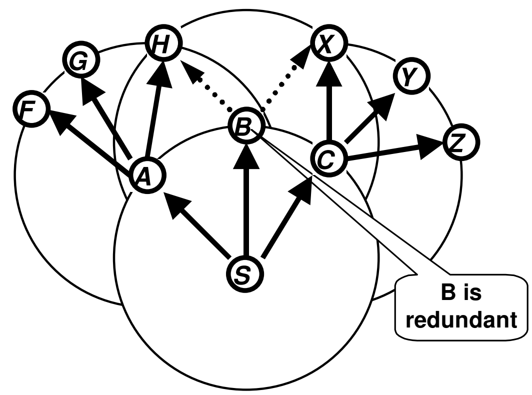

Figure 8: Selective Bordercasting (SBC)

The proposed technique for computing this minimal partitioning set is based on the "greedy" heuristic introduced in [7]. Each of the selected inner peripheral nodes is sent a list (in the route query packet) of the outer peripheral nodes that it partitions. This list becomes the recipient node's set of assigned inner peripheral nodes. The restriction imposed on the set of inner peripheral nodes helps to direct query threads outward from the source, rather than overlap or loop back on themselves.

Unlike the other query control techniques, SBC does not come for free. The amount of IARP traffic increases to provide extended zone topology. In addition, the length of each IERP route query packet increases in order to accommodate the list of assigned peripheral nodes. To be viable, this added cost must be offset by the reduction in query packet transmissions due to overlap prevention.

5. EVALUATION OF THE ZRP

The performance of the ZRP was evaluated based on simulations of 500 node ad-hoc networks, over a range of routing zone radii (ρ), from purely reactive routing (ρ=1 hop) to purely proactive routing ($\rho \rightarrow \infty$ hops). Performance was gauged by measuring the control traffic generated by the ZRP and the average response time of the reactive route discovery process.

[8] Alternatively, IP can unicast the queries if the MAC and IP layers are permitted to operate in promiscuous mode.

[9] For the purpose of this discussion, we will refer to the peripheral nodes of the *routing zone* as *inner* peripheral nodes and the peripheral nodes of the *extended zone* as *outer* peripheral nodes.

Measurements of control traffic is reported in terms of ID fields (rather than packets) transmitted at the network layer. This distinction allows us to account for the variable length of the IERP control packets due to factors such as route accumulation. The overall ZRP control traffic is viewed as the sum of the ID fields in the transmitted intrazone update packets and the interzone route request/reply packets. The neighbor discovery beacons are excluded from our measurement of ZRP control traffic because we assume that this service is already provided in conjunction with the MAC protocol.

The delay performance of the ZRP is reflected by the average delay of an IERP route discovery (delay between the time that a route query packet is issued and the first route response packet is received). Like our measurements of control traffic, we generalize the delay performance by expressing it in terms of the transmission delay of an ID field.

Our simulated network consists of 500 mobile nodes, whose initial positions are chosen from a uniform random distribution over an area of 1500 [m] by 1500 [m]. All nodes move at a constant speed, v, with an initial direction,[10] θ, which is uniformly distributed between 0 and 2π. When a node reaches the edge of the simulation region, it is reflected back into the coverage area, by setting its direction to $-\theta$ (horizontal edges) or $\pi-\theta$ (vertical edges). The magnitude of the velocity is not altered.

For the purposes of our simulation, we assume that there is no MAC layer channel contention. This assumption prevents the ZRP delay measurements from being biased by the delays associated with any particular MAC collision avoidance scheme.

Our assumption of a collision-free media access protocol means that the average SIR of a received packet is limited by the ambient background noise and receiver noise. For fixed transmitter and noise powers, we assume that the BER is reasonably low within a distance, which we call d_{xmit}. Beyond d_{xmit}, the BER increases rapidly. This behavior results from a rapid decrease in received power as the separation distance is increased. We approximate this rapid increase in BER by the following simplified path loss model:

$$PL(d) = \begin{cases} 0 \text{ [dB]} & \text{for } d \le d_{xmit} \\ \infty \text{ [dB]} & \text{for } d > d_{xmit} \end{cases}$$

We interpret this behavior as follows: any packet can be received, error-free, within a radius of d_{xmit} from the transmitter, but is lost beyond d_{xmit}. Since packet delivery is guaranteed to any destination in the range of the source, we are able to further reduce the complexity of our model by eliminating packet retransmission at the data link level.

To accommodate the heavy computational load of simulating a 500 node ad-hoc network, the IARP and IERP are simulated separately. The OPNET™ Network Simulator from MIL3, an event driven simulation package, is used to evaluate the performance of the IARP over a range of routing zone radii. The IARP simulations were run for a duration of 125 seconds. No data was collected for the first 5 seconds of the simulations to avoid measurements during the transient period and to ensure that the initial intrazone route discovery process stabilizes.

[10] Direction is measured as an angle relative to the positive x-axis.

The simulation of the IERP is based on the assumption that the network topology remains constant over the duration of a route discovery[11]. IERP performance measurements are gathered from 2500 route discoveries performed over a total of 50 independent "snapshots" (fixed network configurations) of our network. Each route query is for a destination selected from a uniform random distribution of all nodes outside of the querying node's routing zone. These route queries represent both the initial query performed at the beginning of a session and subsequent queries due to reported route failures.

We assume the average network load to be low. Thus, the queueing delays experienced by route queries are solely due to the bordercasting of a single route query. This assumption is reasonable if the query packets form a separate queue from the actual traffic queue or if they are given a higher transmission priority. We further assume that propagation and node processing delays are negligible.

Parameter	Symbol	Value
Network coverage area	A	1500 [m] x 1500 [m]
Transmission radius	d_{xmit}	100 [m]
Beacon period	T_{beacon}	0.2 [sec]
Transmission rate	R_{xmit}	1.0 [Mbps]

Table 1: Fixed Simulation Parameters

Parameter	Symbol	Values
Routing zone radius	ρ	1-10 [hops]
Node speed	v	10-75 [m/sec]
Mean route query rate	R_{query}	0.1-10.0 [query/s/node]

Table 2: Variable Simulation Parameters

6. PERFORMANCE RESULTS

Results of our simulation are presented in the following figures.

Figure 9 demonstrates the dependence of intrazone control packets on the routing zone radius, ρ, for various rates of network reconfiguration. A distinction is made between the full bordercasting and selective bordercasting schemes because the selective bordercasting requires the IARP to maintain an extended zone of radius $2\cdot\rho$. The increase in IARP traffic resulting from the extended routing zone is shown to be quite significant. In both cases, the amount of IARP control traffic *per node* is approximately proportional to ρ^2. This behavior is to be expected, since the amount of proactive routing traffic *per node* is proportional to the number of nodes that are being "tracked" in the routing zone, and the number of zone nodes is proportional to

[11] The short range radios (d_{xmit} = 0.1 km) are assumed to support transmission rates on the order of at least 100 kbps, resulting in short query transmission delays. This makes our short-term fixed topology assumption reasonable.

the "area" (ρ^2) of the zone. It should be noted that there is no intrazone control overhead for ρ=1. All nodes within a routing zone of ρ=1 are, by definition, neighbors. Consequently, the Neighbor Discovery Protocol provides all of the information needed to maintain connectivity within the routing zone.

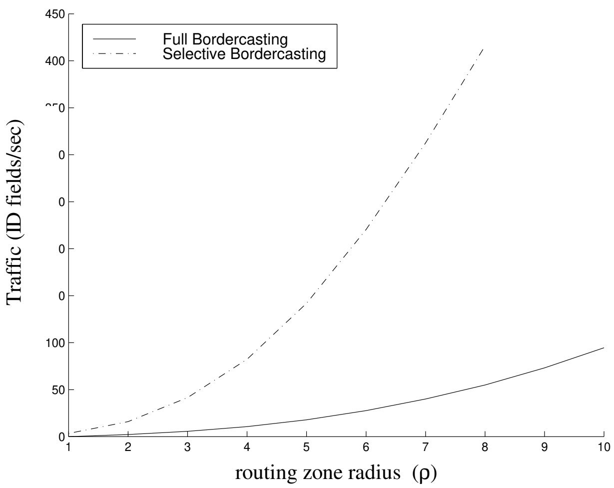

Figure 9: IARP Traffic Per m/s

We examine the behavior of the IERP control traffic by first focusing on the full bordercasting and selective bordercasting cases separately. Figure 10a shows the performance of the query detection/termination techniques that are *effective* in controlling the propagation of IERP traffic in conjunction with full bordercasting. To be considered effective, we require that the amount of IERP traffic per route discovery be a decreasing function of the routing zone radius. We note that some form of advanced query detection (either QD1 or QD2) is needed to properly contain the spread of query packets. Single channel networks, which can implement QD2, may experience approximately 40% less reactive route discovery traffic than those networks that only implement QD1. Early termination (ET), although not "effective" by itself, provides a significant reduction in the amount of IERP traffic when used in combination with the other techniques. As we would expect, the amount of IERP traffic

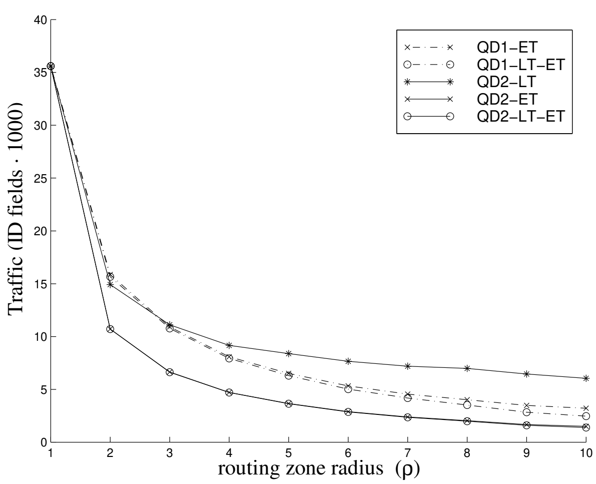

Figure 10a: IERP Traffic Per Route Discovery -- Full Bordercasting

decreases as the query detection capabilities are extended and the termination criteria become stronger. Note the significant boost in performance compared with traditional flooding algorithms (ρ=1).

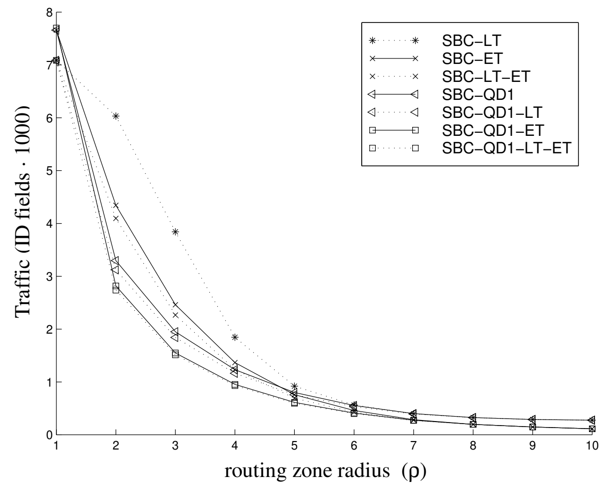

Figure 10b: IERP Traffic Per Route Discovery -- Selective Bordercasting

The performance of selective bordercasting is reflected in Figure 10b. The local overlap prevention provided through selective bordercasting is strong enough to be effective when used in conjunction with *any* combination of LT, ET and QD1. We note the absence of QD2, which proved to be a powerful query control technique when used with full bordercasting. It was discovered that the combination of QD2 and selective bordercasting prevented the IERP route discovery process from achieving full network coverage. This incompatibility occurs because nodes that detect, but do not propagate, *selectively* bordercasted queries, do not necessarily fall under the coverage of the query (due to the focused coverage of a selective bordercast, as compared to the omni-directional coverage of a full bordercast).

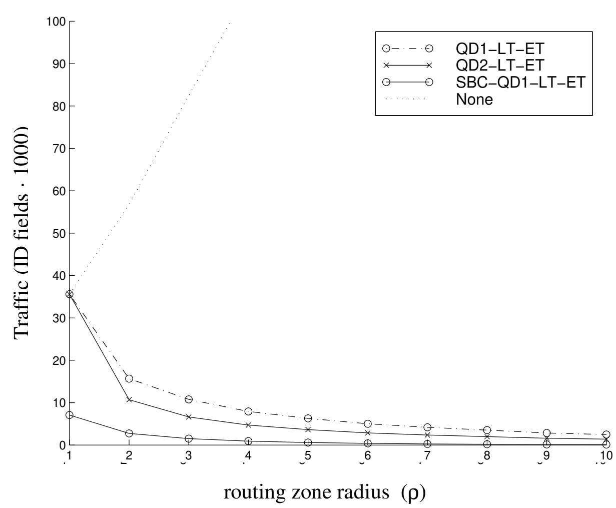

Figure 10c: IERP Traffic Per Route Discovery

Because the LT, ET and QD1/QD2 techniques can be implemented with no additional traffic and negligible computational overhead, the full array of *valid* query control techniques should be applied to provide the best IERP traffic performance. [12]

Figure 10c clearly demonstrates the extent to which the proposed query control mechanisms suppress redundant query traffic. As stated earlier, in the absence of an effective query control strategy, the amount of reactive traffic will increase with the size of the routing zone. When none of the proposed query control schemes are employed, we observe that the amount of query traffic increases linearly with the routing zone radius. For ρ = 3, for example, the IERP without query control generates about twice as much traffic as flooding, and about 10 – 50 times as much traffic as the most effective query control mechanisms.

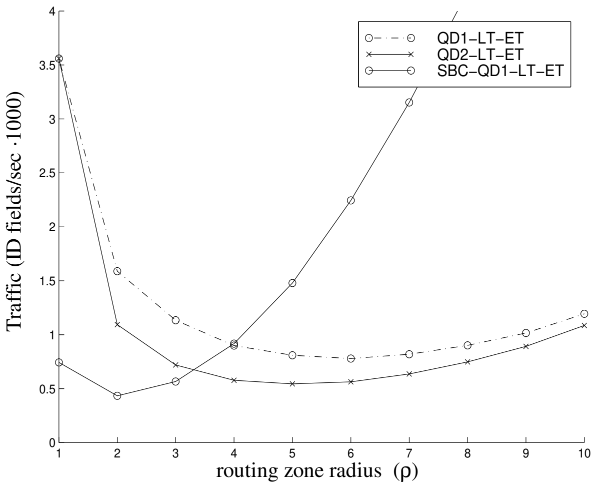

Figure 11a: Total ZRP Traffic
v = 10 [m/sec], R_{query} = 0.1 [query/sec] (CMR=10 [query/km])

Figure 10c also provides a direct comparison between the best IERP traffic performance available from the full and selective bordercasting implementations. All else being equal, selective bordercasting provides a substantial reduction in IERP traffic compared with full bordercasting. In the case of flooding (ρ=1), selective bordercasting generates approximately 20% of the full bordercasting traffic. The impact is even more significant as the routing zone radius increases.

Having analyzed the behavior of the individual IARP and IERP components, we now focus our attention on the total ZRP control traffic. Figures 11 a-i show how the ZRP can be optimized for different conditions of node mobility and call activity, through the adjustment of the routing zone radius. Keeping the route query rate fixed, we see that the optimal routing zone radius decreases as nodal velocity increases. Increased nodal velocity causes the network to reconfigure more rapidly, resulting in an increased of IARP route update traffic. Likewise, we find that, for a constantnode velocity, the optimal routing zone radius increases with the route query rate. Increased call activity results in the generation of additional IERP route query traffic, but has no effect on the reconfiguration rate of the network (i.e. no effect on the IARP traffic). We summarize these trends as follows: increased CMR[13] favors a more proactive ZRP configuration (larger routing zones). Likewise, decreased CMR favors a more reactive ZRP configuration (smaller routing zones).

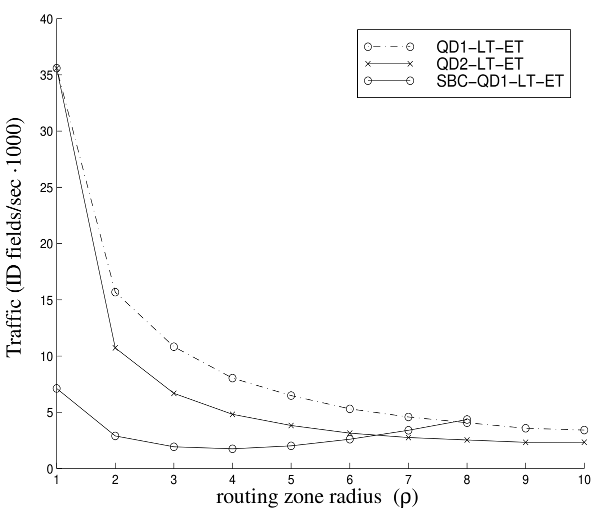

Figure 11b: Total ZRP Traffic
v = 10 [m/sec], R_{query} = 1.0 [query/sec] (CMR=100 [query/km])

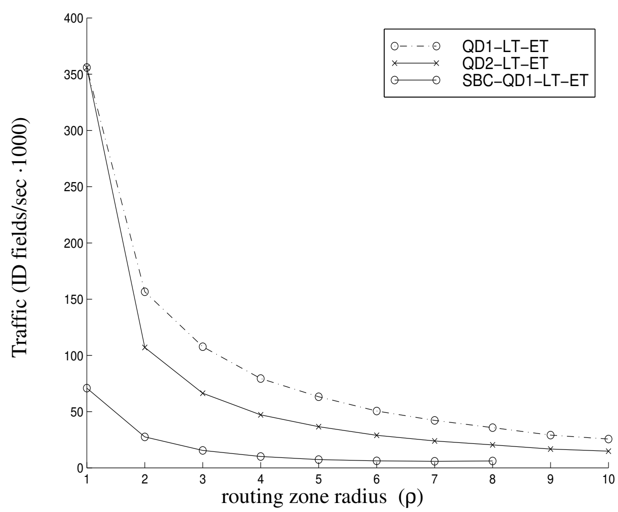

Figure 11c: Total ZRP Traffic
v = 10 [m/sec], R_{query} = 10.0 [query/sec] (CMR=1000 [query/km])

---

[12] Recall that QD2 is not supported by networks which use multiple channels or IERP implementations that use selective bordercasting

[13] Call-to-mobility ratio. Increased CMR corresponds to increased route query rate or decreased node mobility. Likewise, decreased CMR corresponds to either decreased route query rate or increased node mobility.

Comparing selective bordercasting with full bordercasting, we find that the selective bordercasting implementation favors a more reactive ZRP configuration. This is to be expected, since selective bordercasting was shown to improve the efficiency of the reactive IERP, while adding significant cost to the proactive IARP. In single channel networks, the best full bordercasting solution appears comparable to the best selective bordercasting approach. In multi-channel networks, where QD2 may not be employed, selective bordercasting may result in about 50% as much traffic as a full bordercasting approach.

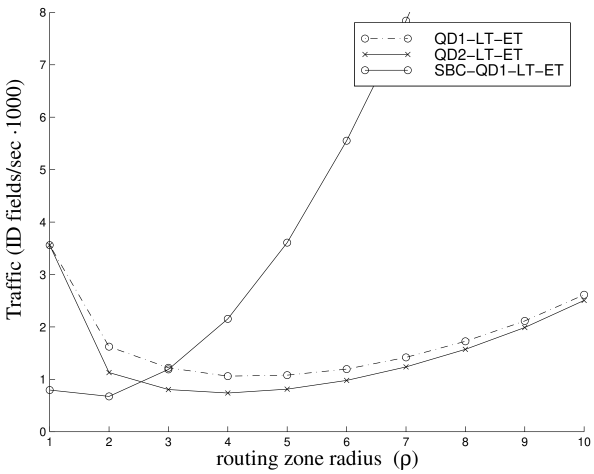

Figure 11d: Total ZRP Traffic
v = 25 [m/sec], R_{query} = 0.1 [query/sec] (CMR=4 [query/km])

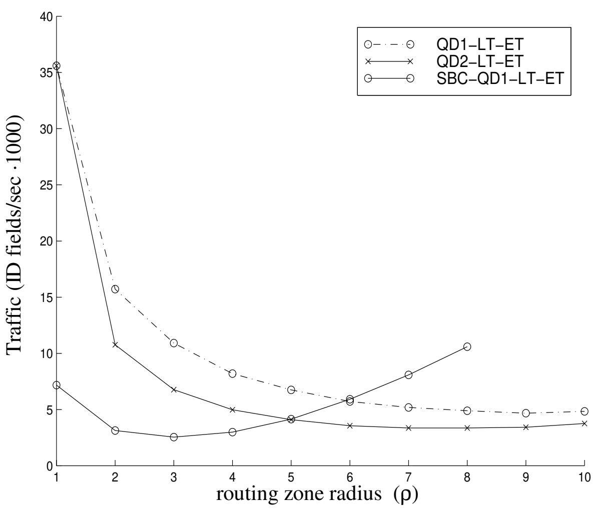

Figure 11e: Total ZRP Traffic
v = 25 [m/sec], R_{query} = 1.0 [query/sec] (CMR=40 [query/km])

We also gauge the performance of the ZRP in terms of route discovery delay. Figure 12 shows that the route discovery delay, like the IERP control traffic, is a decreasing function of the routing zone radius. This relationship is essentially influenced by the same factors that govern the IERP traffic behavior. First, packet transmission time is reduced due to the shorter length of accumulated routes for larger routing zone radii. Second, each query experiences fewer IERP queuing delays due to the increased separation distance between peripheral nodes.

Selective bordercasting schemes exhibit better delay performance than full bordercasting schemes for low routing zone radii, but slightly worse delay performance for larger routing zone radii. At low routing zone radii, selective bordercasting benefits from the reduced queueing delay at each peripheral node. However, at larger radii, the appended list of assigned inner peripheral nodes may be relatively large, resulting in extra transmission delay that offsets the benefits of the reduced queueing delays.

Rather than compare the route discovery delays of full bordercasting and selective bordercasting for the same routing zone radius, a more meaningful comparison is the delay between full bordercasting and selective bordercasting at their respective optimal routing zone radii. Recalling that selective bordercasting operates at a much lower routing zone radius, we see that full bordercasting can respond to a route query in as little as 1/3 the time as selective bordercasting. Given the assumptions behind our delay model, the relative delay performance of selective bordercasting is somewhat optimistic. If the queueing delays due to IARP traffic and the processing delays of the query control algorithms are also factored in, the selective bordercasting can be expected to exhibit relatively worse average route discovery delay.

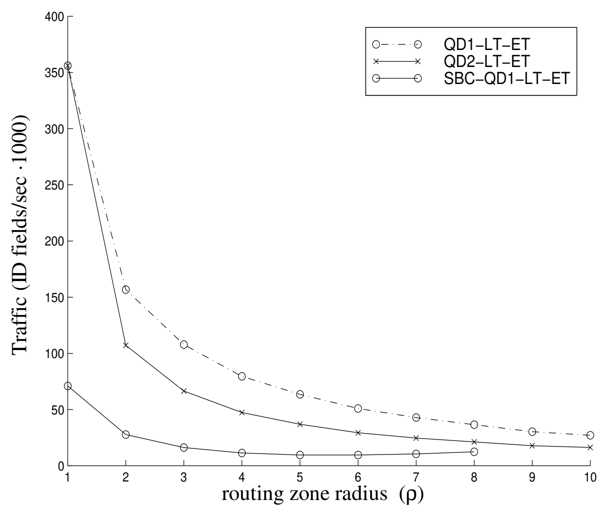

Figure 11f: Total ZRP Control Traffic
v = 25 [m/sec], R_{query} = 10.0 [query/sec] (CMR=400 [query/km])

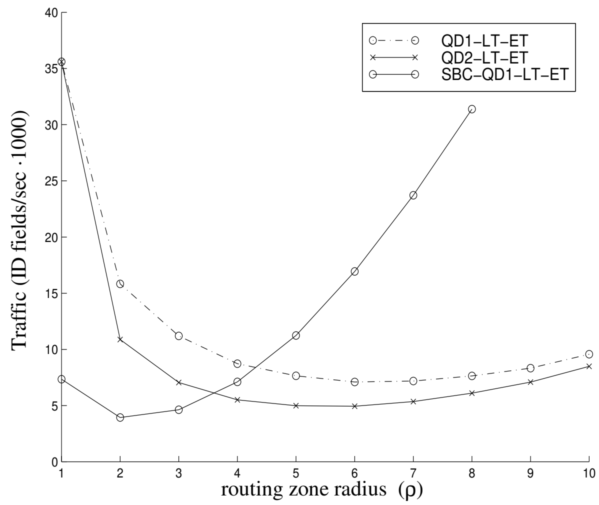

Figure 11g: Total ZRP Traffic
v = 75 [m/sec], R_{query} = 0.1 [query/sec] (CMR=1.3 [query/km])

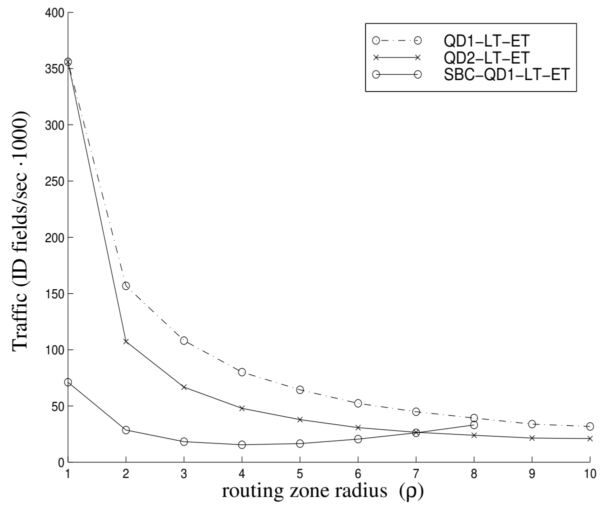

Figure 11i: Total ZRP Control Traffic
v = 75 [m/sec], R_{query} = 10.0 [query/sec] (CMR=130 [query/km])

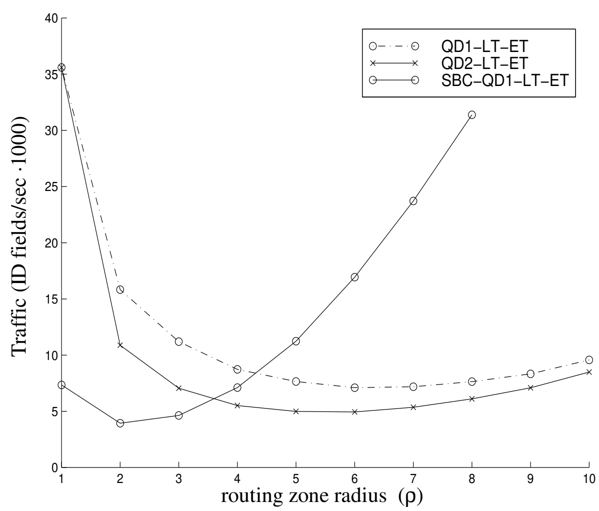

Figure 11h: Total ZRP Control Traffic
v = 75 [m/sec], R_{query} = 1.0 [query/sec] (CMR=13 [query/km])

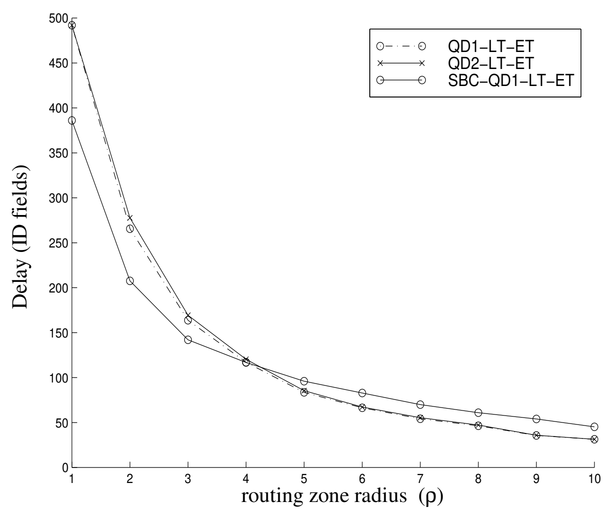

Figure 12: IERP Route Discovery Delay

7. SUMMARY, FUTURE WORK AND CONCLUDING REMARKS

The Zone Routing Protocol (ZRP) provides a flexible solution to the challenge of discovering and maintaining routes in a wide variety of ad-hoc network environments. The ZRP combines two radically different methods of routing into one protocol. Intrazone routing uses a proactive protocol to maintain up-to-date routing information to all nodes within its routing zone. By contrast, interzone route discovery is based on a reactive route request/route reply scheme.

The amount of intrazone control traffic required to maintain a routing zone increases with the size of the routing zone. However, the structure of the routing zone can be exploited to significantly reduce the amount of reactive interzone control traffic. Using a mechanism that we refer to as bordercasting, queries may be passed directly to the periphery of the queried routing zone, without incurring any queueing delays at intermediate nodes. An undesirable side effect of bordercasting is the overlapping of query threads. We have introduced and analyzed the advanced query detection and termination techniques (LT, QD1/QD2, ET) which effectively combat the redundant querying, while generating no additional control traffic and requiring negligible computational overhead. Further reduction of the interzone control traffic can be achieved by preventing thread overlap locally through *selective* bordercasting. Unlike the other query control mechanisms, selective bordercasting requires additional overhead, primarily through the proactive maintenance of an extended zone.

For networks characterized by highly mobile nodes and very unstable routes, the hybrid proactive-reactive routing scheme produces less average total ZRP control traffic than purely reactive ($\rho=1$) or purely proactive ($\rho \rightarrow \infty$) routing. Increasingly reactive ZRP configurations (smaller routing zones) appear to be more suitable for networks that exhibit low call to mobility ratios. On the other hand, networks characterized by slower moving, highly active nodes (frequent route requests), lend themselves to a more proactive configuration (larger routing zones).

Selective bordercasting favors a more reactive ZRP configuration than full bordercasting. For single channel networks, the amount of routing traffic produced through selective bordercasting is comparable to the traffic produced through full bordercasting. For multiple channel networks, however, selective bordercasting produces about half the control traffic as full bordercasting.

We note that for networks with low activity, the instantaneous network load is generally dominated by the control traffic from a single route discovery. Under these conditions, both selective and full bordercasting have been shown to provide noticeably faster route response time than traditional flooding schemes. When the ZRP is configured to minimize total routing control traffic, we find that full bordercasting responds to route queries at least three times faster than a selective bordercasting implementation.

Based on these results, selective bordercasting may be a suitable platform for the IERP in multiple channel networks where conservation of bandwidth is more important than delay performance. In all other cases, it appears that the simpler full bordercasting protocol is the preferred query propagation mechanism.

We have demonstrated that the ZRP may be configured to minimize the amount of routing control traffic, given *a priori* knowledge of the network nodal velocity and route query rate. Recall that the route query rate reflects not only the initial route query for a destination, but also subsequent queries in response to route failure. Thus, the route query rate is not only a function of the communication activity of the node, but is also dependent on node velocity and routing zone radius.

8. ACKNOWLEDGEMENTS

The authors thank the following organizations and individuals for their continuous support of this work:

Dr. Roger Peterson (Motorola), Dr. Michael Kotzin (Motorola), and Mr. Siamak Tabrizi (U.S. Air Force).

9. REFERENCES

[1] Bertsekas, D., and Gallager, R., *Data Networks,* Second Edition, Prentice Hall, Inc., 1992.

[2] Cheng, C., Reley, R., Kumar, S.P.R., and Garcia-Luna-Aceves, J.J., "A Loop-Free Extended Bellman-Ford Routing Protocol without Bouncing Effect," *ACM Computer Communications Review,* vol. 19, no. 4, 1989, pp. 224-236.

[3] Garcia-Luna-Aceves, J.J., "Loop-Free Routing Using Diffusing Computations," *IEEE/ACM Transactions on Networking,* vol. 1, no. 1, February 1993, pp. 130-141.

[4] Haas, Z.J., "A Routing Protocol for the Reconfigurable Wireless Networks," *IEEE ICUPC'97,* San Diego, CA, October 12-16, 1997.

[5] Haas, Z.J., and Pearlman, M.R., "The Zone Routing Protocol (ZRP) for Ad-Hoc Networks," *IETF MANET*, Internet Draft, Dec. 1997.

[6] Johnson, D.B. , and Maltz, D.A. , "Dynamic Source Routing in Ad-Hoc Wireless Networking," in *Mobile Computing,* T. Imielinski and H. Korth, editors, Kluwer Academic Publishing, 1996.

[7] Johnson, D.J., "Approximation Algorithms for Combinatorial Problems," *J. of Computer and System Sciences*, vol. 9, pp. 256-278.

[8] Murthy, S. , and Garcia-Luna-Aceves, J.J., "A Routing Protocol for Packet Radio Networks," Proc. of ACM Mobile Computing and Networking Conference, *MOBICOM'95,* Nov. 14-15, 1995.

[9] Murthy, S., and Garcia-Luna-Aceves, J.J., "An Efficient Routing Protocol for Wireless Networks," *MONET,* vol.1, no.2, pp.183-197, October 1996.

[10] Park, V.D., and Corson, M.S. "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," *IEEE INFOCOM '97*, Kobe, Japan, 1997

[11] Perkins, C.E., and Bhagwat, P., "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," *ACM SIGCOMM,* vol.24, no.4, Oct. 1994, pp.234-244.

[12] Perkins, C.E., "Ad Hoc On-Demand Distance Vector (AODV) Routing,", *IETF MANET,* Internet Draft, Dec.1997.

[13] Tsuchiya, P.F., "The Landmark Hierarchy: A New Hierarchy for Routing in Very Large Networks," *ACM Computer Communications Review*, vol. 18, no. 4, pp.35-42,1988