Abstract

This document describes the Zone Routing Protocol (ZRP) framework, a hybrid routing framework suitable for a wide variety of mobile ad-hoc networks, especially those with large network spans and diverse mobility patterns. Each node proactively maintains routes within a local region (referred to as the routing zone). Knowledge of the routing zone topology is leveraged by the ZRP to improve the efficiency of a globally reactive route query/reply mechanism. The proactive maintenance of routing zones also helps improve the quality of discovered routes, by making them more robust to changes in network topology. The ZRP can be configured for a particular network by proper selection of a single parameter, the routing zone radius.
Contents

Status of this Memo ........................................... i
Abstract ....................................................... i

Applicability Statement ................................. iii
  A. Networking Context ................................. iii
  B. Protocol Characteristics and Mechanisms ........ iii

1. Introduction .................................................. 1

2. Overview of the Zone Routing Framework .......... 2
   2.1 Routing Zones and the Intrazone Routing Protocol ... 2
   2.2 Bordercast-Based Global Reactive (Interzone) Routing . 4

3. The ZRP Architecture ................................. 5

4. Other Considerations ................................. 6
   4.1 Sizing the Routing Zone ......... 6
   4.2 Hierarchical Routing and the ZRP ........ 6

5. Patent Rights Statement .......................... 8

6. References .................................................. 9

Authors’ Information ................................... 11
MANET Contact Information ................................ 11
Applicability Statement

A. Networking Context

The hybrid Zone Routing Protocol (ZRP) framework can adapt to a wide variety of network scenarios by adjusting the range of the nodes’ proactively maintained routing zones. Large routing zones are preferred when demand for routes is high and/or the network consists of many slowly moving nodes. In the extreme case of a network with fixed topology, the ideal routing zone radius would be infinitely large. (Consider the purely proactive routing protocols used on the fixed Internet). On the other hand, smaller routing zones are appropriate in situations where route demand is low and/or the network consists of a small number of nodes that move fast relative to one another. In the "worst case", a routing zone radius of one hop is best, and the ZRP defaults to a traditional reactive flooding protocol.

When the ZRP is properly configured for a particular network scenario, it can perform at least as well as (and often better than) its purely proactive and reactive constituent protocols. In situations where the network behavior varies across different regions, the ZRP performance can be fine-tuned by individual adjustment of each node’s routing zone.

The global reactive component of the ZRP uses the multicast based "bordercast" mechanism to propagate route queries throughout the network efficiently, rather than relying on neighbor-broadcast flooding found in traditional reactive protocols. Consequently, the ZRP provides the most benefit in networks where reliable neighbor broadcasting is either inefficient or altogether impossible. In particular, the ZRP is well suited for multi-channel, multi-technology routing fabrics and networks operating under high load.

B. Protocol Characteristics and Mechanisms

* Does the protocol provide support for unidirectional links? (if so, how?)
  Yes. The ZRP provides "local" support for unidirectional links. A unidirectional link can be used as long as the link source and link destination lie within each other’s routing zone.

* Does the protocol require the use of tunneling? (if so, how?)
  No.

* Does the protocol require using some form of source routing? (if so, how?)
  No. The ZRP’s framework supports global route discovery based on source routing, distributed distance vectors, or various combinations of both.
* Does the protocol require the use of periodic messaging? (if so, how?)

Yes. The ZRP framework proactively maintains local routing information (routing zones) based on periodic exchanges of neighbor discovery messages.

* Does the protocol require the use of reliable or sequenced packet delivery? (if so, how?)

The ZRP only assumes that link-layer (neighbor) unicasts are delivered reliably and in-sequence. Reliable and sequenced delivery of neighbor broadcasts and IP unicasts is not required.

* Does the protocol provide support for routing through a multi-technology routing fabric? (if so, how?)

Yes. It is assumed that each node’s network interface is assigned a unique IP address.

* Does the protocol provide support for multiple hosts per router? (if so, how?)

Yes. Multiple hosts may be associated with a router. These hosts can be identified by the neighbor discovery/maintenance process.

By default, it is assumed that a host’s association with a router is not permanent. As a result, the ZRP exchanges routing information for individual hosts in the same manner as routing information for routers.

In cases where a router is permanently associated with a network (subnetwork), the ZRP supports the exchange of network (subnetwork) prefixes in place of all aggregated host IP addresses.

* Does the protocol support the IP addressing architecture? (if so, how?)

Yes. Each node is assumed to have a unique IP address (or set of unique IP addresses in the case of multiple interfaces). The ZRP references all nodes/interfaces by their IP address.

This version of the ZRP also supports IP network addressing (network prefixes) for routers that provide access to a network of non-router hosts.
* Does the protocol require link or neighbor status sensing
  (if so, how?)

  Yes. The ZRP proactively maintains local routing information
  (routing zones) based on detected changes in neighbor status.

* Does the protocol have dependence on a central entity?
  (if so, how?)

  No. The ZRP is a fully distributed protocol.

* Does the protocol function reactively? (if so, how?)

  Yes. The ZRP’s GLOBAL route discovery mechanism is reactive.
  A route query is initiated, on demand, when a node requires
  routing information that is not immediately available in its
  routing table.

  The route query propagates through the network, using a special
  packet delivery service called "bordercasting". Bordercasting
  leverages knowledge of local network topology to direct route
  queries away from the source, thereby reducing redundancy.

* Does the protocol function proactively? (if so, how?)

  Yes. The ZRP proactively maintains LOCAL routing information
  (routing zones) for each node. The routing zone information is
  leveraged, through the bordercasting mechanism, to support
  efficient global propagation of route queries.

* Does the protocol provide loop-free routing? (if so, how?)

  Yes. If the reactive Interzone Routing is based on source
  routing, loop-freedom in the route discovery process is ensured
  by inspection of accumulated source routes. For distributed
  distance vector approaches, loop-freedom can be ensured by
  labeling queries (replies) with the source (destination) address
  and locally unique sequence number. Nodes that relay these
  messages can temporarily cache these identifiers in order to
  identify and terminate loops.

  The proactive Intrazone Routing based on link states is
  inherently loop-free, although temporary loops may form while new
  link state updates propagate through the routing zone.

* Does the protocol provide for sleep period operation? (if so, how?)

  No. Sleep period operation is not addressed in this draft.
  However, sleep period support can be included as needed.
* Does the protocol provide some form of security? (if so, how?)

   No. It is assumed that security issues are adequately addressed by the underlying protocols (IPsec, for example).

* Does the protocol provide support for utilizing multi-channel, link-layer technologies? (if so, how?)

   Yes. It is assumed that each node’s network interface is assigned a unique IP address.
1. Introduction

One of the major challenges in designing a routing protocol for the ad hoc networks stems from the fact that, on one hand, to determine a packet route, a node needs to know at least the reachability information to its neighbors. On the other hand, in an ad hoc network, the network topology can change quite often. Furthermore, as the number of network nodes can be large, the potential number of destinations is also large, requiring large and frequent exchange of data (e.g., routes, routes updates, or routing tables) among the network nodes. Thus, the amount of update traffic can be quite high. This is in contradiction with the fact that all updates in a wirelessly interconnected ad hoc network travel over the air and, thus, are costly in resources.

In general, the 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. Examples of proactive routing protocols include the Wireless Routing Protocol (WRP) [11] and Destination-Sequence Distance-Vector (DSDV) routing [16]. Reactive protocols, on the other hand, invoke a route determination procedure on demand only. Thus, when a route is needed, some sort of global search procedure is employed. The family of classical flooding algorithms belong to the reactive group. Some other examples of reactive (also called on-demand) ad hoc network routing protocols are Dynamic Source Routing (DSR) [9], Ad-hoc On demand Distance Vector Routing (AODV) [17] and the Temporally Ordered Routing Algorithm (TORA) [13].

The advantage of the proactive schemes is that, once a route is needed, there is little delay until the route is determined. In reactive protocols, because route information may not be available at the time a datagram is received, the delay to determine a route can be quite significant. Furthermore, the global flood-search procedure of the reactive protocols requires significant control traffic. Because of this long delay and excessive control traffic, pure reactive routing protocols may not be applicable to real-time communication. However, pure proactive schemes are likewise not appropriate for the ad hoc networking environment, as they continuously use a large portion of the network capacity to keep the routing information current. Since nodes in an ad hoc network move quite fast, and as the changes may be more frequent than the route requests, most of this routing information is never even used! This results in a further waste of the wireless network capacity. What is needed is a protocol that, on one hand, initiates the route determination procedure on-demand, but at limited search cost. The protocol described in this draft, termed the "Zone Routing Protocol (ZRP)" ([1], [2], [14]), is an example of a such a hybrid proactive/reactive scheme.
The ZRP, on one hand, limits the scope of the proactive procedure only to the node’s local neighborhood. On the other hand, the search throughout the network, although global in nature, is done by efficiently querying selected nodes in the network, as opposed to querying all the network nodes.

A related issue is that of updates in the network topology. For a routing protocol to be efficient, changes in the network topology should have only a local effect. In other words, creation of a new link at one end of the network is an important local event but, most probably, not a significant piece of information at the other end of the network. Globally proactive protocols tend to distribute such topological changes widely in the network, incurring large costs. The ZRP limits propagation of such information to the neighborhood of the change only, thus limiting the cost of topological updates.

2. Overview of the Zone Routing Framework

In the Zone Routing framework, a proactive routing protocol provides a detailed and fresh view of each node’s surrounding local topology (routing zone) at the local level. The knowledge of local topology is used to support services such as proactive route maintenance, unidirectional link discovery and guided message distribution. One particular message distribution service, called bordercasting [5], directs queries throughout the network across overlapping routing zones. Bordercasting is used in place of traditional broadcasting to improve the efficiency of a global reactive routing protocol.

The benefits provided by routing zones, compared with the overhead of proactively tracking routing zone topology, determine the optimal framework configuration. As network conditions change, the framework can be dynamically reconfigured through adjustment of each node’s routing zone.

2.1 Routing Zones and the Intrazone Routing Protocol (IARP)

In Zone Routing, the Intrazone Routing Protocol (IARP) 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 is that the routing zones of neighboring nodes overlap.
An example of a routing zone (for node A) of radius 2 is shown below.

```
+-----------------------------------------+
|                       +---+             |
|            +---+   ---| F |-------|     |
|            | G |-----| D |--/   +---+     +---+   |
| G -----| D |--/  +---+                | E |   | Routing Zone of |
|                  +---+     +---+   |      node A |
|                  |            +---+   ---| B |-------|     |
|                  |            | A |--/   +---+             |
|                  |            |            +---+                        |
+-----------------------------------------+
```

Routing Zone of node A  (radius = 2 hops)

Note that in this example nodes B through F are within the routing zone of A. Node G is outside A’s routing zone. Also note that E can be reached by two paths from A, one with length 2 hops and one with length 3 hops. Since the minimum is less or equal than 2, E is within A’s routing zone.

Peripheral nodes are nodes whose minimum distance to the node in question is equal exactly to the zone radius. Thus, in the above figure, nodes D, F, and E are A’s peripheral nodes.

The construction of a routing zone requires a node to first know who its neighbors are. A neighbor is defined as a node with whom direct (point-to-point) communication can be established 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 separate 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.

Neighbor discovery information is used as a basis for the IARP. IARP can be derived from globally proactive link state routing protocols that provide a complete view of network connectivity. [3] describes the Intrazone Routing Protocol (IARP) in detail and lists the conversion guidelines for adapting a proactive link-state based routing protocol to serve as an IARP.
2.2 Bordercast-Based Global Reactive (Interzone) Routing

Route discovery in the Zone Routing framework is distinguished from standard broadcast-based route discovery through a message distribution service known as the Bordercast Resolution Protocol (BRP) [5]. Rather than blindly broadcasting a route query from neighbor to neighbor, bordercasting allows the query to be directed outward, toward regions of the network (specifically, toward peripheral nodes) that have not yet been "covered" by the query. (A covered node is one that belongs to the routing zone of a node that has received a route query). The query control mechanisms reduce route query traffic by directing query messages outward from the query source and away from covered routing zones.

A node can determine local query coverage by noting the addresses of neighboring nodes that have forwarded the query. In the case of multiple channel networks, a node can only detect query packets that have been directly forwarded to it. For single channel networks, a node may be able to detect any query packet forwarded within the node’s radio range. When a node identifies a query forwarding neighbor, all known members of that neighbor’s routing zone (i.e., those members which belong to both the node’s and neighbor’s routing zones) are marked as covered.

When a node is called upon to relay a bordercast message, it again uses its routing zone topology to construct a bordercast tree, that is rooted at itself and spans its uncovered peripheral nodes. The message is then forwarded to those neighbors in the bordercast tree. By virtue of the fact that this node has forwarded the query, all of its routing zone members are marked as covered. Therefore, a bordercasting node will not forward a query more than once. The details of the Bordercast Resolution Protocol are described in [5].

Given the availability of an underlying bordercast service, the operation of Zone Routing’s global reactive Interzone Routing Protocol (IERP) is quite similar to standard route discovery protocols. An IERP route discovery is initiated when no route is locally available to the destination of an outgoing data packet. The source generates a route query packet, which is uniquely identified by a combination of the source node’s address and request number. The query is then relayed to a subset of neighbors as determined by the bordercast algorithm. Upon receipt of a route query packet, a node checks if the destination lies in its zone or if a valid route to it is available in its route cache. If the answer is affirmative, a route reply is sent back to the source. If not, the node bordercasts the query again. The operation of the IERP is sufficiently general, so that many existing reactive protocols can be used as an IERP with minimal modification. The details of IERP’s operation can be found in [4].
3.0 The ZRP Architecture

The architecture of the hybrid Zone Routing framework is modular, so that a link state routing protocol can be used as an IARP [3] and an on-demand routing protocol can be used as an IERP [4]. As an example, consider TBRPF [12] or OLSR [7] as an IARP and AODV [15] or DSR [8] as an IERP.
4. Other Considerations

4.1 Sizing the Zone Radius

The performance of the Zone Routing Protocol is determined by the routing zone radius. In general, dense networks consisting of a few fast moving nodes favor smaller routing zones. On the other hand, a sparse network of many slowly moving nodes operates more efficiently with a larger zone radius.

The simplest approach to configuring the routing zone radius is to make the assignment once, prior to deploying the network. This can be performed by the network administration, if one exists, or by the manufacturer, as a default value. This may provide acceptable performance, especially in situations where network characteristics do not vary greatly over space and time. Alternatively, the ZRP can adapt to changes in network behavior, through dynamic configuration of the zone radius [6]. In [14], it was shown that a node can accurately estimate its optimal zone radius, on-line, based on local measurements of ZRP traffic. The re-sizing of the routing zone can be carried out by a protocol that conveys the change in zone radius to the members of the routing zone.

In Zone Routing with independently sized routing zones capability, each of the nodes in the network can adaptively configure its own optimal zone radius in a distributed fashion. The performance of Zone Routing is further improved by the ability to provide fine-tuned adaptation to local and temporal variations in network characteristics [19]. The details of the Independent Zone Routing (IZR) framework will be included in a future version of this draft.

4.2 Hierarchical Routing and the ZRP

In a hierarchical network architecture, network nodes are organized into a smaller number of (often disjoint) clusters. This routing hierarchy is maintained by two component routing protocols. An intra-cluster protocol provides routes between nodes of the same cluster, while an inter-cluster protocol operates globally to provide routes between clusters.

The ZRP, with its routing zones and intrazone / interzone routing protocol (IARP/IERP) architecture may give the appearance of being a hierarchical routing protocol. In actuality, the ZRP is a flat routing protocol. Each node maintains its own routing zone, which heavily overlaps with the routing zones of neighboring nodes. Because there is a one-to-one correspondence between nodes and routing zones, the routing zones, unlike hierarchical clusters, do not serve to hide the details of local network topology. As a result, the network-wide interzone routing protocol (IERP) actually determines routes between individual nodes, rather than just between higher level network entities.
For small to moderately sized networks, flat routing protocols, like the ZRP, are preferable to hierarchical routing protocols. In order for a node to be located, its address needs to reflect the node’s location within the network hierarchy (i.e. \{cluster id,node id\}). Because of node mobility, a node’s cluster membership (and thus address) is subject to change. This complicates mobility management, for the benefit of more efficient routing. In many hierarchical routing protocols, each cluster designates a single clusterhead node to relay inter-cluster traffic. These clusterhead nodes become traffic “hot-spots”, potentially resulting in network congestion and single point of failure. Furthermore, restricting cluster access through clusterhead nodes can lead to sub-optimal routes, as potential neighbors in different clusters are prohibited from communicating directly. Some hierarchical routing protocols, such as the Zone-Based Hierarchical Link-State (ZHLS) \[8\], avoid these problems by distributing routing information to all cluster nodes, rather than maintaining a single clusterhead.

In spite of these disadvantages, hierarchical routing protocols are often better suited for very large networks than flat routing protocols. Because hierarchical routing protocols provide global routes to network clusters, rather than individual nodes, routing table storage is more scalable. Similarly, the amount of route update messages is also more scalable. To some extent, the reduction in routing traffic is offset by extra mobility management overhead (i.e. identifying which cluster a node belongs to). However, it is quite common that the environment or existing organizational structure causes nodes to naturally cluster together. In these cases, there may be a high degree of intra-cluster mobility, with inter-cluster mobility is less common.

A hierarchical routing protocol can be viewed as a set of flat routing protocols, each operating at different levels of granularity. In a two-tier routing protocol, the inter-cluster components is essentially a flat routing protocol that computes routes between clusters. Likewise, the intra-cluster component is a flat routing protocol, that computes routes between nodes in each cluster. For example, the Near Term Digital Radio (NTDR) System \[18\] and ZHLS both employ proactive link state protocols to determine inter and intracluster connectivity.

In place of traditional proactive or reactive protocols, we recommend that the intra-cluster and inter-cluster routing protocol components be implemented based on the hybrid proactive/reactive ZRP. As described throughout this draft, the ZRP is designed to provide an optimal balance between purely proactive and reactive routing. This applies equally well to routing between nodes at the intra-cluster level and between clusters at the inter-cluster level. Additionally, the hybrid ZRP methodology can be applied to the related mobility management protocols, which determine complete node addresses based on a node’s location in the network hierarchy.
5. Patent Rights Statement

Cornell University has filed one or more patents protecting the inventions described in this submission (the "Patents"). If Cornell University’s submission, or material portions thereof, is accepted and becomes part of the IETF standard (the "Standard"), then Cornell will grant any entity wishing to practice the Standard a non-exclusive, royalty-free license under the Patents to the extent, but only to the extent, that such license is required to implement (a) mandatory elements of the Standard; or (b) elements of Cornell’s submission that are explicitly specified (and not merely incorporated by reference) in the Standard.

Use of any elements of Cornell’s submission that are neither required in order to implement the Standard nor explicitly specified in the Standard will require an additional license from Cornell University under terms to be negotiated between the parties.

While the protocol disclosed in Cornell’s submission is being evaluated by the IETF, whether on the standards track or otherwise, Cornell University will and hereby does grant any party a license to utilize, test and evaluate such protocol solely for non-commercial, research purposes.
6. References


Authors’ Information

Zygmunt J. Haas
Wireless Networks Laboratory
323 Frank Rhodes Hall
School of Electrical & Computer Engineering
Cornell University
Ithaca, NY 14853
United States of America
tel: (607) 255-3454, fax: (607) 255-9072
Email: haas@ece.cornell.edu

Marc R. Pearlman
389 Frank Rhodes Hall
School of Electrical & Computer Engineering
Cornell University
Ithaca, NY 14853
United States of America
tel: (607) 255-0236, fax: (607) 255-9072
Email: pearlman@ece.cornell.edu

Prince Samar
374 Frank Rhodes Hall
School of Electrical & Computer Engineering
Cornell University
Ithaca, NY 14853
United States of America
tel: (607) 255-9068, fax: (607) 255-9072
Email: samar@ece.cornell.edu

The MANET Working Group contacted through its chairs:

M. Scott Corson
Institute for Systems Research
University of Maryland
College Park, MD 20742
(301) 405-6630
corson@isr.umd.edu

Joseph Macker
Information Technology Division
Naval Research Laboratory
Washington, DC 20375
(202) 767-2001
macker@itd.nrl.navy.mil