

ALTERNATE PATH ROUTING IN MOBILE AD HOC NETWORKS

Marc Pearlman
Zygmunt Haas

Cornell University
School of Electrical Engineering
Ithaca, New York

Peter Sholander
Scientific Research Corporation
Atlanta, Georgia

Siamak S. Tabrizi
ARFL Rome Labs
AFRL/IFGC
Rome, New York

ABSTRACT

This research used a custom event-driven network simulator to evaluate Alternate Path Routing's (APR's) behavior in Mobile Ad Hoc Networks (MANETs). In particular, it investigated how much route-diversity is inherent in typical ad-hoc networks, and the ability of a globally reactive routing-protocol (such as the Zone Routing Protocol) to capture and utilize that route diversity. It then explored the end-to-end capacity improvements provided by APR. Results indicate that APR's benefit is highly dependent on both the network topology and the channel access methods.

INTRODUCTION

Alternate Path Routing (APR) had its origins in the traditional circuit-switched telephone networks, where it reduced call blocking by providing multiple network-routes for the initial call-setup messaging. Examples included Dynamic Nonhierarchical Routing (DNHR) and Dynamic Alternative Routing (DAR) [1]. In contrast to the connection-oriented telephone network, the Internet was designed to provide best-effort *connectionless* data communication. The network-layer Internet Protocol (IP) did not reserve resources on behalf of data streams. Therefore, call blocking was not an issue. If a valid path existed between a data source and destination, data always had the opportunity to be routed through the network. However, a high traffic load could result in large end-to-end delays or packet-buffer overflows.

In circuit-switched networks APR was used to prevent call blocking. Similarly, it could alleviate congestion in IP-based networks by diverting excess traffic to less-loaded network resources. However, until recently, these benefits did not justify the extra storage cost of alternate paths at network routers, and Internet routing has been primarily based on a single least-cost route.

APR in IP Backbone Networks

The recent emergence of IP-based interactive multimedia communication has spurred the development of network services that attempt to offer some Quality of Service (QoS) guarantees. This has led to research on APR in IP networks, which has focused primarily on two key areas: a) construction of alternate route sets; and b) implementation of policies for distributing traffic amongst those multiple routes.

The desired properties of APR-based load balancing are diversity and minimal cost. As such, a variety of algorithms have been developed to address this "k-best path problem" [2], with most schemes focusing on the construction of the least cost set of disjoint paths [3,4]. Unfortunately though, disjoint paths quite often either do not exist or include long paths if they do exist.

After the routing protocol identifies a candidate set of alternate routes, forwarding policies are then needed to control the use of these routes by individual and aggregated IP flows. In circuit-switched networks, one route is typically designated as the primary route, and that route is then used exclusively until it is no longer able to meet the demands of incoming traffic (for example, due to route failure or congestion). At that point, a "crankback" process tries the alternate routes, one-by-one, until it finds a route that can carry the additional traffic. Similarly, IP-

This work was supported by the Air Force Research Laboratory under Contract F30602-00-C-0032

based networks can route a flow along the best available route that satisfies its desired QoS requirements. However, IP networks can also exploit the packetized nature of IP traffic by distributing a flow's traffic among multiple routes [5,6,7,8].

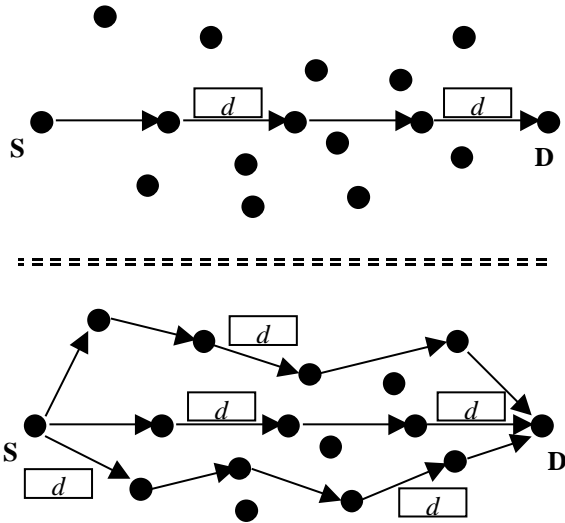


Figure 1. Distributing traffic among multiple routes exploits availability of network resources to increase throughput.

When the primary route is clearly better than the alternate routes, previous research shows that the primary route should be used exclusively until congestion appears, at which time excess traffic can be diverted to secondary route(s). On the other hand, when the maintained routes are of comparable quality, the best strategy is simply to divide the traffic evenly among those routes.

When distributing traffic among multiple routes, a key consideration is the frequency of route switching. From the network-layer perspective, end-to-end throughput and delay improve as the frequency of route transitions increases, with the best policy being to distribute traffic on a per-packet basis [9]. However, alternating routes with such fine granularity can result in out-of-order packet delivery and lead to extra delays for packet re-sequencing at the receiver [10]. Furthermore, out-of-order packet delivery can be misinterpreted by TCP as network congestion, which causes TCP to reduce its effective transmission rate. (A companion paper discusses the effects of APR-induced out-of-order packet delivery on throughput in MANETs [11]. Results indicate that it is an acceptable impairment.)

APR in MANETS

APR's ability to provide load balancing and enhanced survivability makes it an attractive technique for bandwidth-limited MANETs [12] that are designed as packet-radio extensions to the wired Internet. However, the APR performance gains achieved on the wired Internet

do not necessarily carry over to MANETs. In particular, the overlapping radio-coverage of neighboring nodes can result in strong interdependence between alternate routes which limits APR's benefits to particular MANET topologies and channel access techniques.

Current Internet routing protocols, such as Open Shortest Path First (OSPF) and Routing Information Protocol (RIP), were developed for the fixed Internet backbone where routers are bolted to the ground. It is well-known that both protocols simply fail in the mobile wireless environment because they attempt to continuously track every change in the network topology. In wireless environments, this "proactive" approach to routing causes wasted overhead that often saturates the wireless medium with control traffic for routes that are never used. In contrast, mobile wireless networks require reactive routing-protocols that only track routes between source/destination pairs that are actively communicating. The Zone Routing Protocol (ZRP) is one such reactive protocol that provides a routing framework that is well-suited for source-based alternate path routing. ZRP actually uses a mix of reactive and proactive routing [13]. It uses proactive routing, such as traditional link-state routing, within each node's local neighborhood, or "routing zone", and reactive on-demand routing for communications with non-local destinations. As such, a link state version of the ZRP's IntraZone Routing Protocol (IARP) can provide each node with up-to-date connectivity (and optionally, link/node QoS metrics) within its routing zone. For local destinations that lie within the routing zone, a pre-computed set of alternate paths are immediately available when needed. In contrast to the proactive IARP, the reactive Interzone Routing Protocol (IERP) provides routes, as needed, for destinations that lie outside of the routing zone. When the network layer cannot relay a packet because a route to the destination is not known, a route-query for that destination is initiated. If a route exists, that query will result in a set of responses, each explicitly defining a unique path to the destination. These discovered paths can then be directly entered into the list of APR candidate routes, without any extra route-computation. Source routing can then be used to identify the chosen APR path as individual packets are forwarded through the network.

The ZRP Route-Reply messages effectively provide a partial snapshot of the network topology. From the perspective of APR route-set construction, the most value can be obtained from the route-replies by decomposing the reported routes into a collection of links. The links returned from one route-query can be pooled with valid links from other route-queries, and the proactively-tracked links from within the source node's local routing zone, in

order to maximize the effective routing information (see Figure 2a-c).

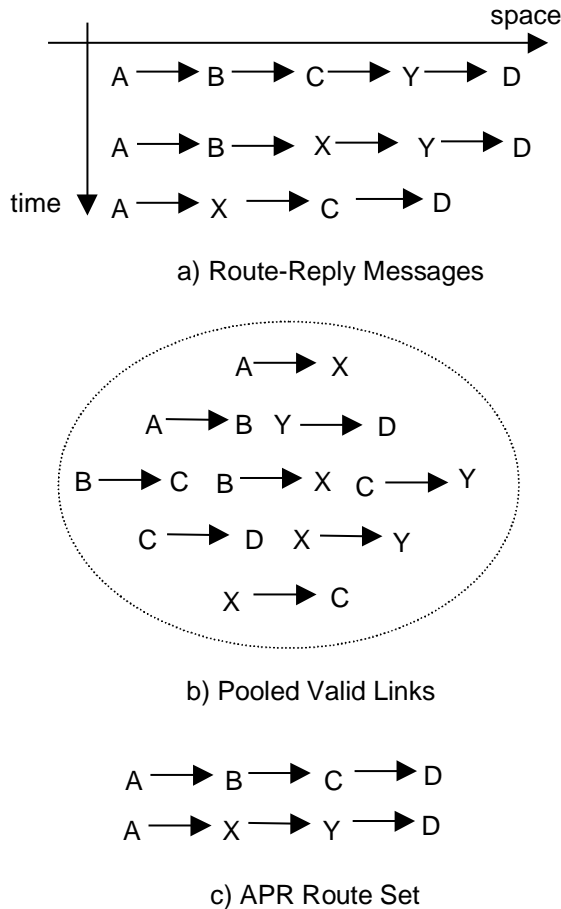


Figure 2. Decomposing Route Replies into Links States in order to Reconstruct Shorter, More Diverse APR Routes

Over an IP flow’s lifetime, the status of the APR routes will probably change in both wired and wireless networks. In particular, as the MANET’s topology evolves, links will be created and destroyed, resulting in route reduction or route failure. Additionally, changes in network traffic (including the traffic introduced by the source itself) will have a significant effect on traffic-dependent route metrics. This motivates the need for a route-probing mechanism that could periodically re-assess the status of the APR route set. The results of those route probes could then be used to both construct a new APR route set and also adjust the route alternation policies. However, this paper will not consider route-probing further.

EVENT-DRIVEN SIMULATION MODEL

This research effort used a custom event-driven network simulator to evaluate APR’s behavior in MANETs. In particular, it investigated how much route-diversity is inherent in typical ad-hoc networks, and the ability of a reactive routing-protocol to capture and utilize that route diversity. It then explored APR’s effects on end-to-end

capacity in MANETs. A companion paper discusses other performance metrics such as delay, packet loss and service interruptions [11].

Mobility Model

The simulation uses an ad hoc network of 100 mobile nodes, whose initial positions are chosen from a uniform random distribution over an area of 1000 m by 1000 m. Each node j moves at a constant speed, v , and is assigned a new direction (as measured as an angle relative to the +X-axis), θ_j , 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 $-\theta_j$ (horizontal edges) or to $\pi-\theta_j$ (vertical edges). The magnitude of its velocity is not altered.

Physical Layer Model

In the absence of a packet collision, the simulation model assumes that a packet can be delivered (error-free) to any receiver within d_{xmit} from the transmitting node. Receivers farther than d_{xmit} from the transmitting node will not receive that packet. (This simple threshold packet-delivery model is a reasonable approximation for digital transmission in the presence of background channel-interference and receiver noise.) This paper’s results assume $d_{xmit} = 140$ meters, which corresponds to an average of six neighbors per (non-boundary) node. The user data-rate of each channel is 100 kbps.

Channel Access Model

The simulator models both “single-channel” and “multiple-channel” multi-access methods. In single-channel, multiple-access networks, nodes contend for the channel based on a channel access protocol. This presents two novel problems not found in more familiar broadcast Ethernet channels – namely “hidden nodes” and “exposed nodes”. A “hidden node” occurs when both Nodes A and C want to transmit to Node B. If Node B can hear both A and C, but C can not hear A then collisions (which waste precious wireless-network bandwidth) can occur at Node B. In contrast, an “exposed node” occurs when Node B can hear Nodes A and C, but Nodes A and C are out of radio propagation range of each other. In that case, Node B may needlessly defer transmission to Node C if Node A is active.

This simulator uses the Dual Busy Tone Multiple Access (DBTMA) protocol [14], wherein a node secures access to the channel through an RTS/CTS handshake (performed possibly on a separate control channel) prior to transmitting a data packet. After completing the RTS/CTS handshake, the transmitter sends the data packet, while simultaneously activating a “transmit busy-tone”. The intended receiver, in turn, activates a separate “receive busy-tone” as soon as this data transmission is detected.

The dual busy tones are used to block attempts by neighboring nodes to access a channel already in use. In particular, the transmit busy-tone prevents the transmitter’s neighbors from accepting incoming RTS requests. Likewise, the receive busy-tone prevents the receiver’s neighbors from initiating the RTS/CTS handshake. These busy-tones effectively prevent the “hidden terminal problem”. In addition, DBTMA inherently avoids the “exposed terminal problem”, by permitting neighboring nodes to transmit data simultaneously to different (and available) receivers.

Single-channel networks are simple to operate but have several drawbacks. For example, in single channel networks, while a transmitter (Tx) is sending a packet to a neighboring receiver (Rx), Tx’s other neighbors are blocked from receiving data from other sources.

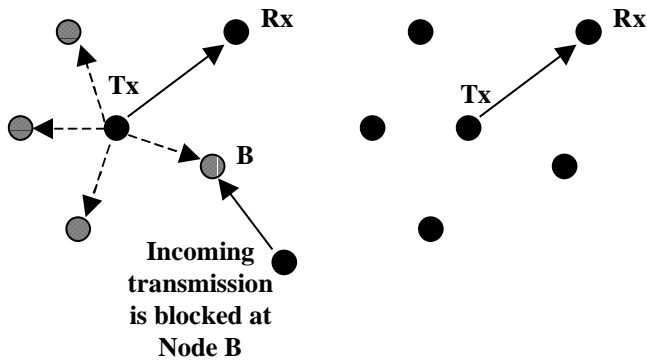


Figure 3. Single Channel versus Multiple Channel Access Models

In contrast to the single-channel networks, the simulation model assumes contention-free channel access for its multiple-channel. In that case, data transmission from Tx to Rx does *not* block Tx’s other neighbors from receiving data from other sources. The underlying media access control is responsible for assigning each incoming/outgoing link a locally unique channel (frequency, time slot, code,...) to avoid channel contention. Although there are no packet collisions, link-layer retransmissions are still possible, as a receiving node may be busy receiving or transmitting another packet. In addition, since each neighbor receives data on a locally unique channel, data transmission from Tx to Rx does *not* block Tx’s other neighbors from receiving data from other sources.

Neighbor Discovery

The simulation model assumes that neighbor discovery is based on the reception of “HELLO beacons” that are broadcast at the MAC layer. These short beacons (containing only the source address) are transmitted at random intervals of mean T_{beacon} seconds, and are given highest transmission priority. The size of the link-layer

HELLO beacon-packets is 10 bytes. If a new beacon fails to arrive within a pre-defined time of the most recent beacon, then a link failure is reported to the ZRP routing process. Finally, because the simulated links are bi-directional, the need for a more complex HELLO→I-HEAR-YOU packet exchange is eliminated.

Routing Protocol

The simulation model uses the Zone Routing Protocol (ZRP). All nodes use a common routing-zone radius that minimizes the overhead caused by the routing-control traffic for each scenario’s network operating conditions. Finally, the APR algorithm uses the set of all locally-available IARP and IERP information in order to compute a k best-path set of routes to the destination node. For these simulations, k is limited to 1 or 2 routes.

SIMULATION RESULTS

The simulation effort examined the effect of alternate path routing (APR) on two distinct classes of traffic. The first traffic class was unreliable data streams that ran over UDP. Those data streams were characterized by their average throughput and burstiness. Second, the interaction between APR and TCP congestion control was explored based on observations of 1 Mbyte reliable file-transfers.

For each simulation scenario, a single TCP or UDP test session was selected for analysis. For simplicity, the simulation concentrated on source - destination nodes separated by a minimum distance of six hops, as this was approximately the average distance to a randomly chosen destination in this network. The remainder of the network traffic was assumed to be heterogeneous and diverse in nature. Rather than model these other data flows explicitly, the simulation represented this *background* traffic as the load relayed by nodes to their neighbors. For simplicity, the simulation assumed that the background traffic arrived at a node’s network layer (independent of other nodes) according to a Poisson arrival-process, and that each arriving packet was forwarded to a randomly chosen neighbor. Finally, all data traffic (test APR traffic and background traffic) was carried in fixed length packets of 1024 bytes. (Note: the TCP acknowledgement packets were 40 bytes in length, while the link-layer HELLO beacons were only 10 bytes.)

The APR test session was not initiated during the first minute of the simulation, in order to give the background traffic and intrazone route-discovery processes sufficient time to stabilize. Fifty independent simulations were run for each scenario.

Channel-Coupling Effects

Ad-hoc networks may be implemented over a single, shared broadcast channel, or multiple channels (for

example, receiver-oriented codes in spread spectrum communication). The underlying channel can have a significant impact on APR performance, due to “route coupling”. Two routes that have nodes or links in common are considered highly coupled. However, route coupling may occur even if two routes have no common nodes or links. In the case of multiple-channel spread-spectrum networks, transmission across a link may result in degraded quality for a simultaneous transmission on a neighboring link. In single-channel networks, a transmission can block transmission across neighboring links.

Figure 4 demonstrates the problem of route coupling in ad-hoc networks by focusing on route pairs of equal length, with a minimum length of 6 hops. In single-channel networks, each node can expect to find about two nodes in the other routes that share the same channel. In contrast, multiple-channel routes exhibit less than 25% the amount of route coupling. Another observation is that the most diverse routes become less coupled as their lengths increase. This makes sense because longer routes can achieve greater spatial (and hence radio coverage) separation.

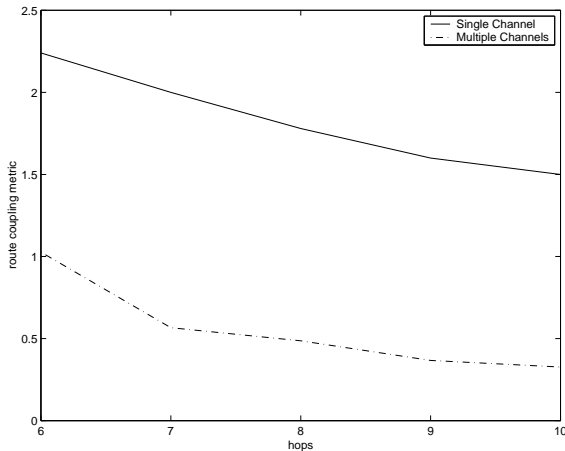


Figure 4. Coupling Between the Most Diverse APR Routes

The next subsection shows that there is an intuitive relation between route coupling and alternate path routing. Load balancing becomes more effective as route coupling decreases and the routes operate independently. On the other hand, if the routes are strongly coupled, traffic on one route may block traffic on alternate routes, preventing performance gains.

End-To-End Capacity Improvements

The maximum achievable end-to-end throughput in MANETs is determined by the interaction of neighboring links. For example, if all nodes have a single half-duplex transceiver then they cannot successfully transmit to a

neighbor while that neighbor is transmitting. As a result, the maximum throughput, for half-duplex transceivers, across a multi-hop route in a multiple-channel network is only half the link data-rate.

For shared-channel networks, a node may be prevented from receiving a packet due to the transmission of its downstream neighbor. As shown in Figure 5, Node B can not receive a packet from Node A, while it is relaying a packet to Node C. Similarly, Node B is blocked from receiving a packet from Node A, while Node C is transmitting on the shared channel. These effects limit the route capacity to one-third of the link data rate, for routes of three hops or more.

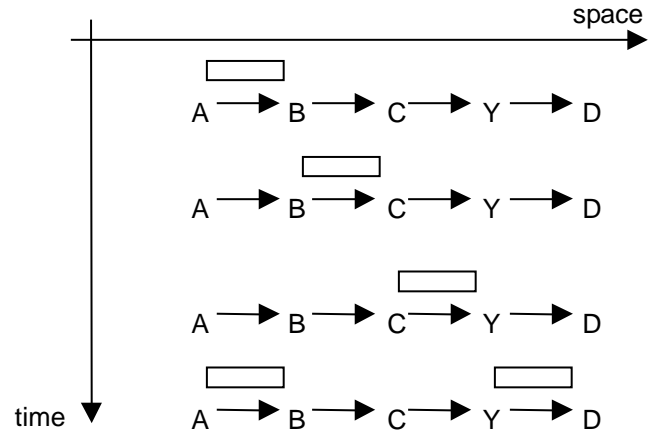


Figure 5. Maximum route capacity for single-channel networks is one-third the link capacity

Given two independent routes, alternate path routing can be used to increase the end-to-end throughput. For multiple-channel networks, alternating between two independent and comparable routes can (in theory) provide a maximum route capacity equal to the link capacity. For shared-channel networks, two independent routes can double the end-to-end throughput. A third route could then allow maximum end-to-end throughput to be achieved.

Practical considerations often limit APR’s ability to provide this increased end-to-end capacity. The set of candidate routes, typically acquired by reactive routing protocols such as ZRP, often have links in common. This problem can be addressed, to some extent, through techniques like the previously discussed link-based route construction. However, even when the route links are unique, the proximity of the nodes in even totally-disjoint alternate routes can produce a coupling effect in shared-channel networks.

Figure 6 shows the end-to-end capacity for typical source-destination pairs separated by at least six hops in the simulated MANET. In multiple-channel networks, a

single route (no APR) will provide 50 kbps (again, half the 100 kbps link capacity). Evenly distributing traffic among two routes can provide extra capacity. Since more diversity can be achieved with larger routes, it is not surprising that larger routes provide greater additional bandwidth. However, APR provides less benefit in single-channel ad-hoc network for at least two reasons. First, the routes in single-channel networks are highly coupled, which leads to a negligible increase in throughput when APR is applied. Secondly, while seven hop routes provide more bandwidth than the minimum-distance six hop paths, capacity is actually lost when longer routes are applied. This can be attributed to the meandering of longer routes, which induces an additional coupling effect on each route.

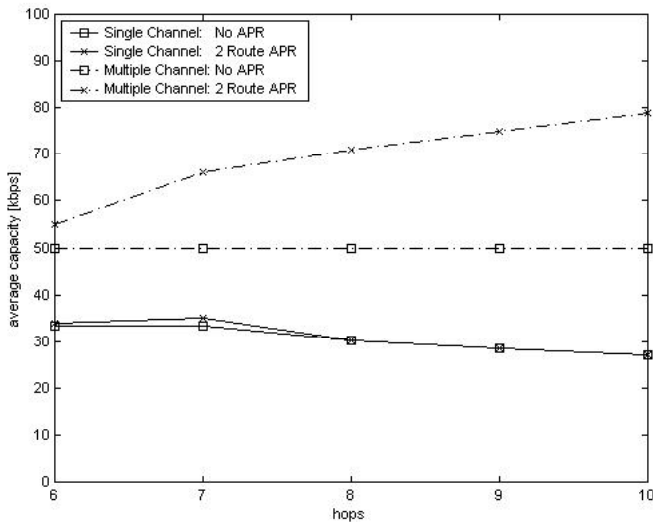


Figure 6. APR Capacity Increase in Ad Hoc Network

CONCLUSIONS AND RECOMMENDATIONS

Alternate path routing (APR) has been applied to telephone networks, ATM and the Internet to support load balancing and survivability. The potential benefits of APR make it appear to be an ideal candidate for the bandwidth limited and dynamic mobile ad-hoc networks (MANETs). This paper's investigation of APR in the MANET environment has revealed that APR can, in some circumstances, provide notable improvements to end-to-end capacity. Quite often, however, the network topology and channel characteristics limit what APR is able to achieve. Whereas these factors can be addressed in wired networks through the deliberate addition and rearrangement of routers and cables, this level of administration contradicts the distributed and self-organizing philosophy of MANETs. This does not mean that APR cannot flourish in this environment. Indeed, proper design choices, such as multiple channels instead of a single broadcast channel, can help to realize some of APR's potential. Other features, like multiple full-duplex

transceivers per node and tighter integration of routing with media access control and power control may extend APR's abilities even further.

REFERENCES

- [1] R.J. Gibbens, F.P. Kelly and P.B. Key, *Dynamic Alternative Routing, Routing in Communications Networks*, ed. by M. Steenstrup, Prentice Hall, 1995.
- [2] S-W Lee and C-S Wu, "A k-Best Paths Algorithm for Highly Reliable Communication Networks", *IEICE Transactions on Communications*, p. 586-590, 4/99.
- [3] R. Ogier and N. Shacham, "A Distributed Algorithm for Finding Shortest Pairs of Disjoint Paths", *IEEE INFOCOM '89*.
- [4] D. Sidhu, R. Nair and S. Abdallah, *Finding Disjoint Paths in Networks*, ACM SIGCOMM '91.
- [5] S. Bahk and M. El Zarki, "Dynamic Multi-path Routing and How it Compares with other Dynamic Routing Algorithms for High Speed Wide Area Networks", ACM SIGCOMM '92, pp. 53-64, 8/92.
- [6] S. Sibal and A. DeSimone, "Controlling Alternate Routing in General-Mesh Packet Flow Networks", ACM SIGCOMM '94, August 1994.
- [7] F. Borgonovo, "Deflection Routing", *Routing in Communications Networks*, edited by M. Steenstrup, Prentice Hall, 1995.
- [8] D. Zappala, "Alternate Path Routing for Multicast", *IEEE INFOCOM 2000*, March 2000.
- [9] R. Krishnan and J. Silvester, "Choice of Allocation Granularity in Multipath Source Routing Schemes", *IEEE INFOCOM '93*, pp. 322-329, March 1993.
- [10] N. Gogate and S. Panwar, "Assigning Customers to Two Parallel Servers with Resequencing", *IEEE Transactions on Communications Letters*, p. 119, 4/99.
- [11] M. Pearlman, Z.J. Haas, P. Sholander and S.S. Tabrizi, work in preparation for invited paper in *IEEE Communications Magazine*.
- [12] H. Siraj, "On the Relationship Between Route Selection Policies and Route Demand in Ad-Hoc Networks", M. Eng. Design Project Report, Cornell University, May 1999.
- [13] "Zone Routing Protocol", see Internet Draft at <http://www.ietf.org/html.charters/manet-charter.html>
- [14] Z.J. Haas and J. Deng, "Dual Busy Tone Multiple Access (DBTMA): A Medium Access Control for Multihop Networks," *WCNC '99*, New Orleans, LA, Sept. 1999.