VI. CONCLUSION

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). Our investigation of APR in the MANET environment has revealed that APR can, in some circumstances, provide as much as a 40% improvements in end-to-end traffic delay. Quite often, however, the network topology and channel characteristics severely 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. Proper design choices can help to realize some of APR's potential. Multiple channel systems provide better link/route isolation than single broadcast channel networks. Routing protocols that provide a thorough view of network connectivity can fully reveal whatever diversity exists in the network. 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, edited 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, vol. E82-B, no. 4, p. 586-590, April 1999.

[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] N. Taft-Plotkin, B. Bellur and R. Ogier, *Quality-of-Service Routing Using Maximally Disjoint Paths*, IEEE IWQoS '99, June 1999.

[6] S. Sibal and A. DeSimone, *Controlling Alternate Routing in General-Mesh Packet Flow Networks*, ACM SIGCOMM"94, August 1994.

[7] R.G. Gallager, *A Minimum Delay Routing Algorithm Using Distributed Computation*, IEEE Transactions on Communication, vol. 25, pp. 73-84, Jan. 1977.

[8] J.J. Garcia-Luna-Aceves, S. Vutukury and W.T. Zaumen, *A Practical Approach to Minimizing Delays in Internet Routing*, IEEE ICC'99, June 1999.

[9] F. Borgonovo, *Deflection Routing*, Routing in Communications Networks, edited by M. Steenstrup, Prentice Hall, 1995.

[10] Z. Wang and J. Crowcroft, Shortest Path First with Emergency Exits, ACM SIGCOMM'90, Philadelphia, PA, Sept. 1990.

[11] R. Krishnan and J. Silvester, *Choice of Allocation Granularity in Multipath Source Routing Schemes*, IEEE INFOCOM'93, pp. 322-329, March 1993.

[12] N. Gogate and S. Panwar, *Assigning Customers to Two Parallel Servers with Resequencing*, IEEE Transactions on Communications Letters, vol. 3, num. 4, p. 119, April 1999.

[13] N. Gogate and S.S. Panwar, *Supporting Applications in a Mobile Multihop Radio Environment Using Route Diversity, I. Non-Real Time Data*, IEEE ICC'98, Atlanta, GA, June 1998.

[14] N. Gogate, D. Chung, S. Panwar, Y. Wang, *Supporting Image/Video Applications in a Mobile Multihop Radio Environment Using Route Diversity*, IEEE ICC'99, Vancouver, BC, June, 1999.

[15] N.F. Maxemchuk, *Dispersity Routing*, IEEE ICC'75, San Francisco, CA, June 1975.

[16] E. Ayanoglu, I. Chih-Lin, R.Gitlin, J. Mazo, *Diversity Coding for Self-Healing and Fault Tolerant Communication Networks*, IEEE Trans. on Communication, vol. COM-41, pp. 1677-1688, Nov. 1993.

[17] 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.

[18] Nasipuri, A. and Das, S.R., "On-Demand Multipath Routing for Mobile Ad Hoc Networks," IEEE ICCCN, Boston, MA, October, 1999.

[19] Moy, J., "OSPF version 2," IETF RFC 2328, April 1998.

[20] 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.

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

[22] Perkins, C.E. and Royer, E.M., “Ad Hoc On-Demand Distance Vector Routing,” *IEEE WMCSA’99*, New Orleans, LA, Feb. 1999.

[23] Pearlman, M.R. and Haas, Z.J., “Determining the Optimal Configuration of the Zone Routing Protocol,” *IEEE JSAC*, vol. 17, num. 6, Aug. 1999.

[24] M. Pearlman and Z. Haas, *Improving the Performance of Query-Based Routing Protocols Through ' Diversity Injection*,' WCNC'99, New Orleans, LA, Sept. 1999.

[25] 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.

The diversity offered by alternate path routing may contribute to improved capacity and reduced delays. However, the use of multiple diverse routes also increases a session's exposure to interruption by link failure. This tradeoff is a particular concern in the MANET environment, where node mobility and limited transmission power conspire to make link failures a frequent occurrence.

Figure 8 illustrates the tradeoff between diversity and survivability, for our ad-hoc network with a uniform node speed of 10 [m/s]. Whereas the median lifetime for a single six hop route is 2.2 seconds, a typical pair of independent six hop routes will last only one second before experiencing a failure. In practice, the drop in interruption times resulting from APR may be less significant, as route coupling and limited knowledge of the network topology make completely independent route sets somewhat of a rarity. Of course, coding techniques (i.e. minimum descriptions coding, diversity/dispersity coding) can be applied to protect the user from service interruption. However, such protection requires extra bandwidth and may counteract any performance gains that might have been achieved through load balancing.

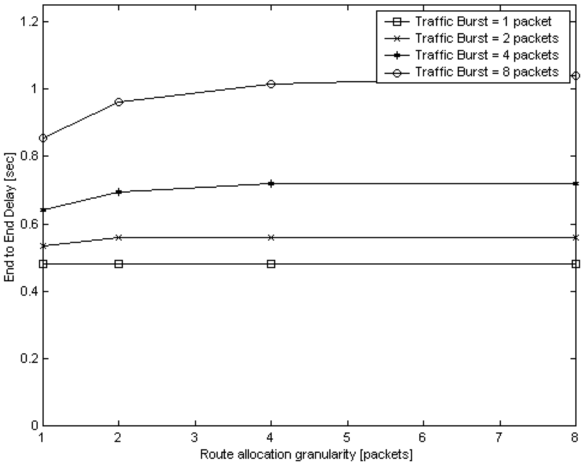

Figure 6c: average end-to-end delay of 16 kbps UDP session
multiple channels, proactive routing,
low network load (no background traffic)

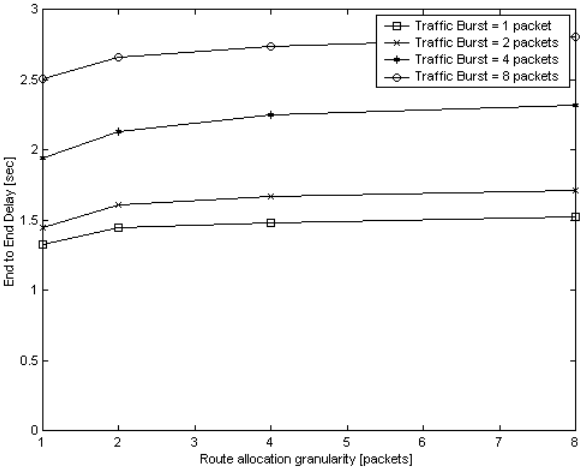

Figure 6d: average end-to-end delay of 16 kbps UDP session
multiple channels, proactive routing,
high network load (5 kbps/node/channel)

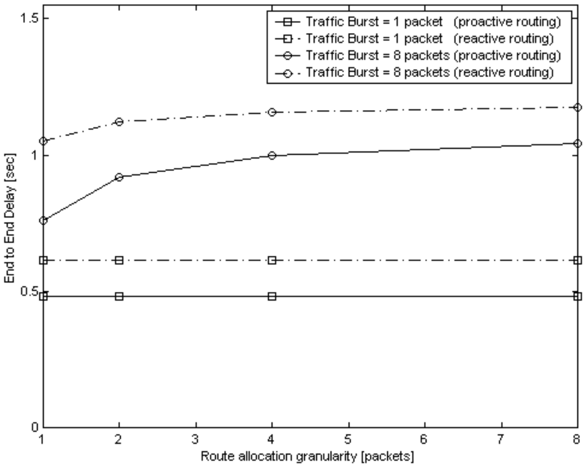

Figure 7: average end-to-end delay of 16 kbps UDP session
multiple channels, low network load (5 kbps/node/channel)

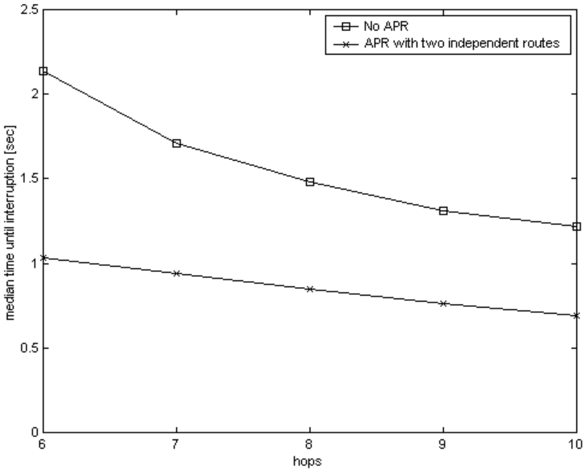

Figure 8: tradeoff between route diversity and session stability

become less coupled as their lengths increase. This makes sense because longer routes can achieve greater spatial (and hence radio coverage) separation.

We now turn our attention to end-to-end delay performance. As shown in Figures 6a-6d, the greatest reduction in delay occurs for bursty packet sources. The batch arrival of a packet burst results in queueing delays as the packets wait to be transmitted by the source. By distributing the packets among multiple routes, this backlog can be reduced faster. Furthermore, load balancing becomes more effective with finer granularity in route allocation. In particular, we note that alternating between routes on a per packet basis makes the best use of available network resources. In contrast, when the route allocation granularity is at least as large as the traffic burst size, APR essentially behaves as a single route scheme.

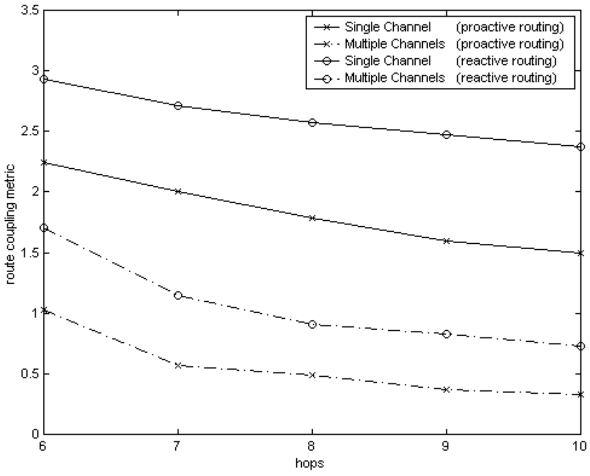

Figure 5: coupling between most diverse APR routes

On *average*, the improvements in end-to-end delay are somewhat modest. For multiple channel networks, delay can be cut by an average of 20% for bursty sources. For single channel networks, the improvements in delay are negligible. The limited gain is again the result of route coupling. The most effective load balancing occurs over independent routes. We've seen that independent minimum distance route pairs are rare. When they do occur, delay may be reduced by up to 40%. More diversity can be found with longer routes, but longer routes also introduce transmission delays over the additional hops. In our simulations with six hop source-destination pairs, we've found that greatest reduction in delay generally occurs when traffic is distributed over one six hop and one seven hop route. This combination strikes the best balance between diversity and length.

As expected, APR has less impact on delay performance when used in conjunction with reactive routing protocols. To illustrate, we consider the multiple channel network under low load operation (see figure 7). We begin by observing that reactive routing protocols lead to more end-to-end delay than proactive routing protocols. This is because a node may not have acquired sufficient connectivity information to construct shortest hop (or nearly shortest hop) paths. Each extra hop accumulates additional transmission and queueing delays. Moreover, the limited diversity between route pairs affects the improvements in delay relative to "no APR". This is demonstrated by the bursty traffic source (burst size of 8 packets). In this case, reactive routing provides only half the relative delay improvement as compared to proactive routing (compare per-packet APR (1 packet granularity) with "no APR" (8 packet granularity)).

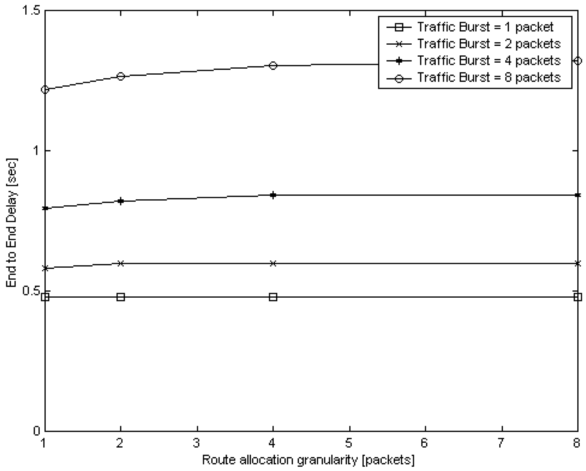

Figure 6a: average end-to-end delay of 16 kbps UDP session
single channel, proactive routing,
low network load (no background traffic)

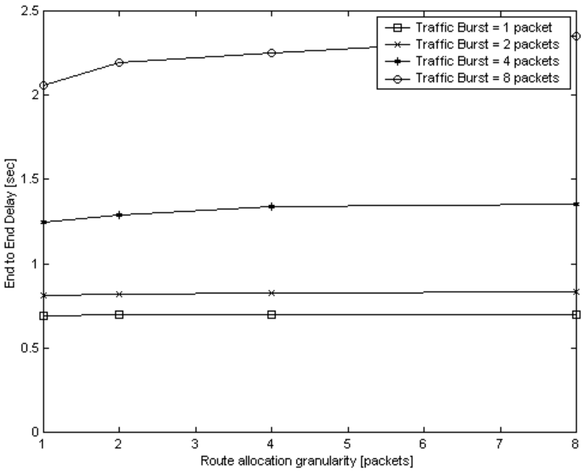

Figure 6b: average end-to-end delay of 16 kbps UDP session
single channel, proactive routing,
high network load (5 kbps /node/channel)

In the absence of a packet collision, we assume that background channel interference and receiver noise limit the transmission range of packets and busy tones to a physical radius of d_{xmit} = 140 [m]. Within the range of d_{xmit}, the average power of the desired signal (and resulting average SIR) rapidly increase to support reliable packet transmission. As significant improvements can be realized through the addition of error control coding, we approximate the rapid increase of packet reliability by a simple threshold packet delivery model: *Once access to the channel has been established*, 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 the packet correctly. This transmission radius corresponds to an average of six neighbors per (non-boundary) node.

In our model, neighbor discovery is based on the reception of HELLO beacons that are broadcast at the MAC layer. These short beacons (containing only source address) are transmitted at random intervals of mean T_{beacon} = 0.2 [sec]. Neighbor connectivity is determined by the reception of the HELLO beacons. If a new beacon fails to arrive within $2 \cdot T_{beacon}$ of the most recent beacon, a link failure is reported. In our simplified ad-hoc network environment, links are bi-directional, eliminating the need for a more complex HELLO→I-HEAR-YOU packet exchange[3]. Furthermore, we assume that neighbor discovery packets are given highest transmission priority and are not destroyed by collisions. This prevents inaccurate reporting of link failures in the $2 \cdot T_{beacon}$ window.

The hybrid proactive / reactive Zone Routing Protocol (ZRP) is used to evaluate network connectivity. Each node tracks the topology of its routing zone with a link-state IARP. Routes for destinations beyond a node's routing zone are acquired, as needed[4], through the IERP's global route discovery procedure. When a node receives a route query packet, it records in a temporary query cache[5] (1) the ID of the packet's previous hop and (2) the hop count from the query source. Route replies are issued in response to the first received query packet only if a valid route to the query destination is locally available. Diversity injection is used to construct a route reply path that promotes increased diversity of the discovered routes. In particular, a node relays a route reply through the cached previous hop that has been selected least often and is closest to the query source. As the reply progresses back to the source node, the reply packet accumulates the IDs of the relaying nodes. When the reply packet reaches the source node, the explicit path is extracted and decomposed into a set of links. The set of all locally available link state information is used to compute the set of k APR paths that minimize the level of APR route coupling (see Section II). In these simulations, we limit k to 1or 2 routes. We present APR performance results for the extreme ZRP configurations of zone radius 1 hop (purely reactive) and ∞ hop (purely proactive).

In this paper, we examine the effect of alternate path routing on unreliable data streams running over UDP. The data streams are characterized by their average throughput and burstiness. For this kind of traffic, we are interested in the maximum supported throughput, end-to-end delay and packet loss rate. For each simulation scenario, a single UDP test session is selected for analysis. For simplicity, we focus our attention on source - destination nodes separated by a minimum distance of six hops, as this is approximately the average distance to a randomly chosen destination in this network. The remainder of the network traffic is assumed to be heterogeneous and diverse in nature. Rather than model these data flows explicitly, we represent this *background* traffic as the load relayed by nodes to their neighbors. For simplicity, we assume that packets to be relayed arrive at a node's network layer (independent of other nodes) according to a Poisson process and are forwarded to a randomly chosen neighbor. All data traffic (test APR traffic and background traffic) is carried in fixed length packets of 1 [Kbyte]. The APR test session is 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.

V. SIMULATION RESULTS

Figure 5 demonstrates the problem of route coupling in ad-hoc networks by focusing on route pairs of equal length. Complete knowledge of the network topology (in this case, by means of a purely proactive ZRP configuration), reveals the best APR paths that exist in the network. For the best routes available in single channel systems, a transmitting node in one path is expected to prevent two nodes from receiving data along the other route. Multiple channel networks exhibit about one third as much route coupling, due to locally unique channel assignments. With a more limited view of network connectivity (as provided by most reactive protocols), the selection of possible APR candidate paths is also limited. For multiple channels, the link state snap-shot provided by the route query produces APR route pairs that are twice as coupled as the network's best routes. Single channel systems exhibit similar behavior. Finally, we observe that the most diverse routes

---

[3] Note that wireless links are not necessarily bi-directional, due to differences in node transmission power, receiver sensitivity, co-channel interference levels, etc. In general, the I-HEAR-YOU response is necessary to verify that a link is, in fact, bi-directional.

[4] In particular, a route discovery is initiated when the network layer (eg. IP) is unable to forward a packet, due to the lack of a route to the packet's destination.

[5] This information is recorded for *all* route query packets, including those that are discarded.

Reactive routing protocols *could* report a broader view of the network by allowing a node to respond to a route query more than once. Alternatively, we can apply "diversity injection" [24] to increase diversity without generating additional route replies. During the route query stage, nodes temporarily cache (1) the previous hop node ID and (2) hop count of *all* received query packets (including the packets that are discarded). Later, during the reply phase, the cached path information is used to redirect replies along more diverse paths back to the source. In particular, replies are relayed through the shortest of the least selected cached paths that does not create a reply loop.

Once collected, route replies can provide the most value when they are decomposed into their constituent links. The links returned from one route query can be pooled with valid links from other route queries to construct a new set of routes with improved diversity and reduced length (Figure 4).

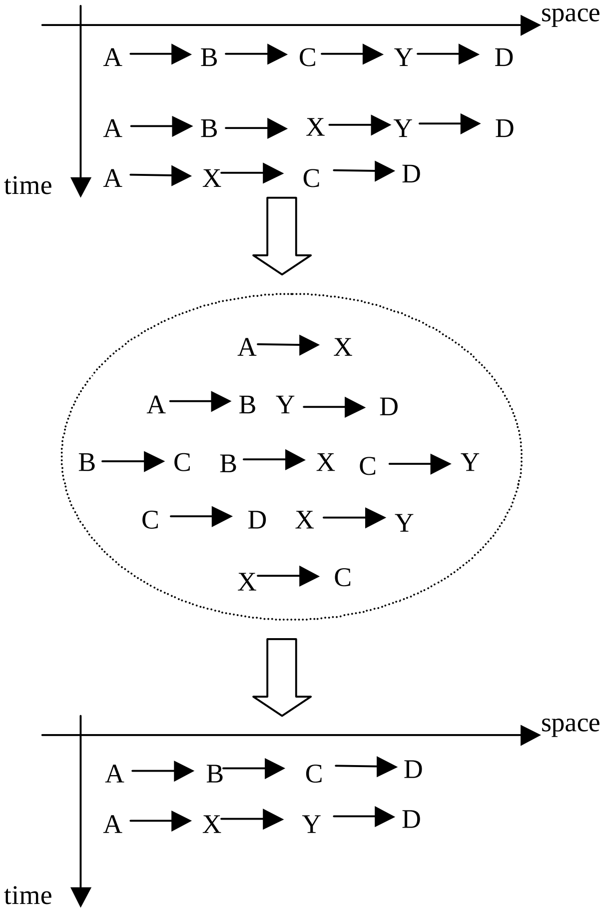

Figure 4: Decomposing route replies into link states to reconstruct shorter, more diverse APR routes

IV. SIMULATION MODEL

We have used a custom event driven network simulator to evaluate the behavior of alternate path routing (APR) in the mobile ad-hoc network environment. In particular, we have investigated how much diversity is inherent in the ad-hoc network, and the ability of the underlying routing protocol to reflect this diversity. We have then explored the relationship between APR in the ad-hoc network and critical performance metrics such as end-to-end capacity, delay, packet loss and service interruptions.

Our ad-hoc network consists 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[2], 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 the velocity is not altered.

In single channel networks, nodes contend for the channel based on a channel access protocol, for example the Dual Busy Tone Multiple Access (DBTMA) protocol [25]. Prior to transmitting a data packet, a node secures access to the channel through an RTS/CTS handshake (performed possibly on a separate control channel). 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 neighbors of the transmitter from accepting incoming RTS requests. Likewise, the receive busy tone prevents the receiver's neighbors from initiating the RTS/CTS handshake. This effectively prevents the "hidden terminal problem" associated with wireless channel access. In addition, DBTMA inherently avoids the "exposed terminal problem", by permitting neighboring nodes to transmit data simultaneously to different (and available) receivers.

In contrast to the single channel networks, we assume that channel access in our multiple channel networks is contention free. 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, retransmissions are still possible, as a receiving node may be busy receiving or transmitting another packet.

---

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

this case, a data transmission does not block the transmitters other neighbors from receiving data from other sources.

The underlying channel scheme can have a significant impact on the performance of APR, 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 nodes or links in common. In the case of multiple channel spread spectrum networks, packet transmission may result in degraded quality of a simultaneous transmission on a neighboring link. In single channel networks, a node's transmission can prevent neighbors from receiving separate transmissions altogether.

We gauge the coupling between two routes (route$_1$ and route$_2$) as the average number of nodes that are unable to receive data along route$_2$ when a single node in route$_1$ is transmitting. First, consider a node that is common to both routes. While it is relaying data along route$_1$, it is unable to receive or transmit[1] data along route$_2$ (because each node is equipped with only a single half-duplex transceiver). Next, consider a node$_1$ in route$_1$ whose route$_1$ downstream neighbor belongs to both routes. When node$_1$ relays route$_1$ data to this neighbor, that neighbor is unable to receive or transmit data along route$_2$. In single channel networks, node$_1$'s transmission will prevent all neighboring nodes (not just the route$_1$ downstream neighbor) from receiving data along route$_2$.

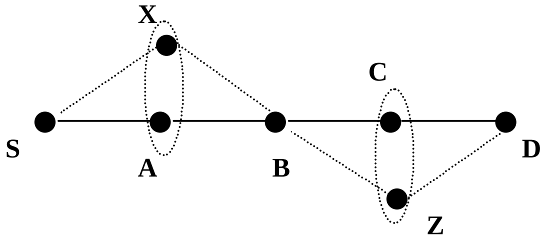

Figure 3a: Coupling of two nearby routes
route$_1$ = S→A→B→C→D
route$_2$ = S→X→B→Z→D

As a brief example, we measure the route coupling between the two four-hop routes illustrated in figures 3a and 3b. When node S transmits to node A (route$_1$), it cannot simultaneously relay data along route$_2$ to node X. Node A's transmission to node B prevents node B from receiving data from node X and relaying data to node Z along route$_2$. Furthermore, in single channel networks, neighboring node X is also blocked by A's transmission. Node B's transmission to node C blocks the same set of nodes (B,Z and in single channels, X). Finally, when C relays data to D, D is blocked from receiving data along route$_2$. In the case of a shared channel, node C's transmission also blocks simultaneous route$_2$ data reception for neighbors B and Z.

In this example, a transmitting node in route$_1$ will block an average of 1.50 nodes in route$_2$ from data reception, in a multiple channel environment. In a single channel environment, 2.50 route$_2$ nodes are blocked, a 67% increase in route coupling.

Transmitting route$_1$ node	Blocked route$_2$ node *multiple channel*	Blocked route$_2$ node *single channel*
S	X	X
A	B, Z	X, B, Z
B	B, Z	X, B, Z
C	D	B, Z, D
Total blocked nodes	6	10
Avg. blocked nodes	1.50	2.50

Figure 3b: Summary of route coupling for example in figure 3a

As we will see later, 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.

III. EXPANDING THE REPORTING CAPABILITIES OF REACTIVE ROUTING PROTOCOLS

The quality of an APR route set depends on the amount of information each node has about the network topology. Proactive routing protocols, like the link-state OSPF [19], provide each node with complete and up-to-date view of the network connectivity. Equipped with this complete information, a node is capable identify the best APR routes that exist in the network. However, in the presence of sufficient node mobility, the tracking *all* changes in network connectivity can become prohibitively expensive. Recent developments in MANET routing protocols have focussed on globally reactive (on-demand) route discovery [20,21,22,23]. When a node needs a route, it issues a route query packet that is relayed through the network. The route query may produce multiple responses containing paths to the sought-after destination. For most reactive routing protocols, route responses provide only a partial snap-shot of network connectivity and therefore typically do not report the "best" routes that the network has to offer.

---

[1] When a node is unable to transmit along a route, it follows that the node's downstream neighbor is blocked from receiving along that route.

The packetized nature of Internet traffic can be exploited by distributing a session's traffic among multiple routes. The optimal division of packet data streams among multiple paths, to minimize routing delay, was described in [7,8]. In high-speed packet networks, network congestion is generally a temporary and local phenomena. Under these circumstances, a reasonable APR strategy would be to direct traffic along a single shortest hop route, bypassing temporarily congested areas when necessary. This philosophy is reflected in schemes like "deflection routing" [9] and "shortest path first with emergency exits" [10].

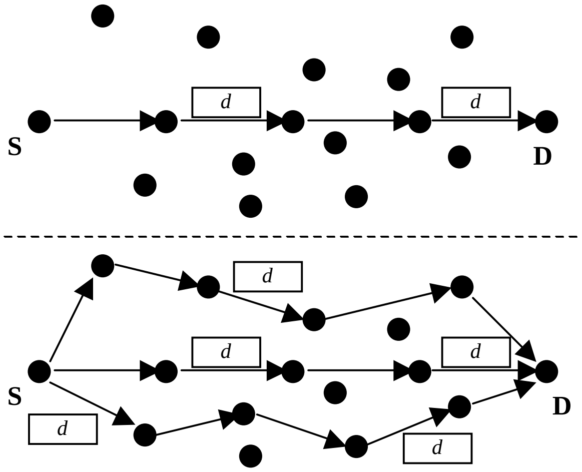

Figure 1: Distributing traffic among multiple routes exploits availability of network resources to increase throughput and reduce delay.

When distributing traffic among multiple routes, the frequency of route switching is a key consideration. 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 [11] (see Figure 1). However, alternating routes with such fine granularity can result in out-of-order packet delivery, leading to extra packet resequencing delays [12]. Furthermore, out-of-order packet delivery can be misinterpreted by TCP as network congestion [13].

Although congestion control has been the primary focus of APR research, the potential of APR to compensate for route failures has also been investigated by the networking community. Multimedia applications can take advantage of multiple routes through multiple descriptions coding [14]. Such coding schemes distribute information among multiple outgoing streams so that reduced quality content can still be provided in the presence of route failure. In addition, "diversity/dispersity" coding [15,16] can be performed in conjunction with APR at the network layer (see figure 2). Error control coding across multiple paths can provide perfect data recovery in the presence of a limited number of route failures.

The potential for APR to address load balancing and survivability makes it an attractive technique for the bandwidth limited and mobile ad-hoc networks [17,18]. Because the current generation of Mobile Ad-Hoc Networks (MANETs) is being designed as a packet radio extension to the Internet, MANETs should be able to support APR schemes developed for general IP networks. However, this does not mean that the APR performance gains achieved on the wired Internet necessarily carry over to MANETs. In particular, overlapping radio coverage of neighboring nodes can result in strong interdependence between alternate routes.

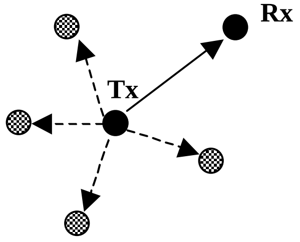

Figure 2a: single channel scheme

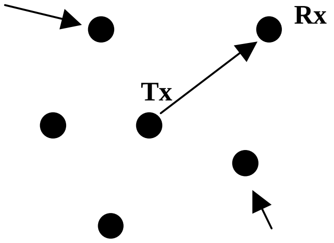

Figure 2b: multiple channel scheme

II. ROUTE COUPLING IN AD-HOC NETWORKS

Ad-hoc networks may be implemented over a single channel (figure 2a), or multiple channels (figure 2b). In single channel systems, nodes transmit and receive on the same shared channel. Neighboring transmitters contend for channel access by means of a media access control (MAC) protocol. When a transmitter gains control of the channel, it can proceed to transmit data to its target neighbor receiver(s). Because all nodes listen on the same channel, the transmission will arrive at all the transmitter's neighbors. Thus, these nodes are unable to receive data from other sources at the same time. In order to support concurrent data transmission in overlapping neighborhoods, multiple channels need to be employed. In particular, we may assign a locally unique channel (frequency, time slot, code) to each receiver or link. In

On the Impact of Alternate Path Routing for Load Balancing in Mobile Ad Hoc Networks

Marc R. Pearlman and Zygmunt J. Haas
School of Electrical Engineering
Cornell University
Ithaca, NY 14853
E-mail: {pearlman,haas}@ee.cornell.edu

Peter Sholander
Scientific Research Corporation
Atlanta, GA 30339
E-mail: psholand@scires.com

Siamak S. Tabrizi
Air Force Rome Laboratories
Rome, NY 13441
E-mail: siamak.tabrizi@rl.af.mil

*Abstract-***Alternate path routing (APR) can provide load balancing and route failure protection by distributing traffic among a set of diverse paths. These benefits make APR appear to be an ideal candidate for the bandwidth limited and mobile ad-hoc networks. However, we find that APR's potential is not fully realized in ad-hoc networks because of route coupling resulting from the geographic proximity of candidate paths between common endpoints. In multiple channel networks, coupling occurs when paths share common intermediate nodes. The coupling problem is much more serious in single channel networks, where coupling also occurs where one path crosses the radio coverage area of another path. The network's inherent route coupling is further aggravated by the routing protocol, which may provide an incomplete view of current network connectivity.**

Through analysis and simulation, we demonstrate the impact of route coupling on APR's delay performance in ad-hoc networks. In multiple channel environments, APR is able to provide a 20% reduction in end-to-end delay for bursty data streams. Though these gains are appreciable, they are about half what we would expect from APR with independently operating routes. Route coupling is so severe in single channel networks that APR provides only negligible improvements in quality of service.

I. INTRODUCTION

The notion of alternate path routing (APR) has its origins in the traditional circuit switched telephone networks. The telephone network core consists of a fully connected set of switches. Under "normal" conditions, a call may be switched across the core in a single hop. Should this primary path become unavailable (due to a link failure or full capacity), a call may be blocked. The desire to avoid call blocking led to alternate routing schemes such as Dynamic Nonhierarchical Routing (DNHR) and Dynamic Alternative Routing (DAR) [1]. In addition to the primary, one-hop path, each switch maintains a set of costlier two-hop paths, which may be used to bypass the primary path if necessary.

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 overflow. Whereas alternate path routing was used to prevent call blocking in telephone networks, it could alleviate Internet congestion by diverting excess traffic to less loaded network resources. In practice, these benefits did not justify the extra storage cost of alternate paths at network routers. As a result, Internet routing has been primarily based on a single, least cost, route.

The emergence of interactive multimedia communication on the Internet has spurred the development of network services to improve quality of service (QoS) and even attempt to offer some QoS guarantees. This has led to a number of approaches for alternate path routing. Research in this area has focused primarily on two key areas: construction of alternate route sets and implementation of policies for traffic distribution among these multiple routes. The desired properties of load balancing with alternate routes are diversity and minimal cost. A variety of algorithms have been developed to address this "k-best path problem" [2]. Most schemes focus on constructing the least cost set of disjoint paths [3,4]. Quite often, however, disjoint paths do not exist, or if they do, may include long paths. Placing constraints on the path lengths can produce a set of paths that have acceptable cost, with a minimal amount of overlap [5].

With a set of candidate alternate routes in hand, policies are needed to control the use of these routes. In circuit-switched data networks, it is customary to designate one route as the primary route. The primary route is used exclusively until it is no longer able to meet the demands of incoming traffic (for example, due to route failure or congestion). Through a *crankback* process, the alternate routes are tried, one-by-one, until a route is identified that satisfies the additional traffic load. This *call packing* approach reduces the fragmentation of network resources that would prevent new connections from being established (due to insufficient bandwidth along one path). Call packing can be used to reduce blocking probabilities in virtual circuit networks as well [6]