A NETWORKING PERSPECTIVE OF MOBILE PARALLEL RELAYS

Yingbo Hua, Yu Chang, Yan Mei

Department of Electrical Engineering
University of California
Riverside, CA 92521, USA
yhua@ee.ucr.edu

ABSTRACT[1]

For mobile ad hoc networking, relaying is one of the most fundamental functions. The traditional relaying strategy is serial where data packets hop from a single node to another, and the nodes neighboring a transmitting node or a receiving node are suppressed from transmission to avoid radio interference. In this paper, we consider a parallel relaying strategy where neighboring nodes may act as parallel relays with space-time modulation. We demonstrate the feasibility of parallel relaying in ad hoc networks by presenting a generic route discovery algorithm for establishing routes of parallel relays. We also formulate a link layer protocol for forwarding packets through parallel relays. We then provide analytical results of packet loss rate and link delay time to highlight some of the potential benefits of parallel relays. With N parallel relays at each link, a diversity factor N-squared is achievable. A power saving of 10 *dB* or more is possible from serial relaying to parallel relaying.

1. INTRODUCTION

For mobile ad hoc networks, the traditional relaying strategy [1-2] is serial where a data packet is transferred from a single node to another. The serial relaying strategy is highly vulnerable to small scale fading that is inherent in radio signals propagating over terrain. For outdoors and non-line-of-sight transmission, the applicable wavelength may become too large to use an antenna array on a handheld device. To increase the robustness against small scale fading, it is desirable for multiple neighboring nodes to perform parallel relaying.

With the parallel relaying strategy, a data packet may be simultaneously transferred from multiple nodes to a single node [4-7]. In this paper, we consider regenerative relays where a relaying node forwards a packet only when the packet was received correctly. We will refer to a relaying node simply as relay. We assume that the mobile wireless parallel relays at each link can be synchronized with each other at the symbol level[2]. Therefore, in the transmission mode, the parallel relays may apply the space-time modulation techniques [4-7, 9].

Space-time modulation is a technique to achieve an efficient averaging of diverse channel fading factors. It ensures a good performance without the need of channel feedback. The alternative approach of switching to the best node at all time would add a significant burden on networking control especially in a highly mobile environment. Other cooperative methods are available via [8].

In section 2, we present a generic route discovery algorithm to establish a route of parallel relays in an arbitrary setting of mobile nodes. In section 3, we show how data packets can be relayed through parallel relays. In section 4, we give analytical and numerical results to highlight some of the potential benefits of using parallel relays. In particular, we show that the diversity factor of a chain of N parallel relays at each tier can be as high as N^2. The power saving from serial relays to parallel relays can be as high as 10 *dB* or more.

2. FINDING A ROUTE OF PARALLEL RELAYS

Research in ad hoc networking has been exclusively limited to serial relays [1-2]. Finding a route of parallel relays does not appear to be an obvious task. To develop a route discovery algorithm for parallel relays, we use and generalize an existing framework for serial relays. Specifically, we assume that each node in the network has a unique identification (ID) and a route table. The route table lists the *next-hop* neighboring nodes for each potential destination node in the network. A link quality factor (e.g., packet loss rate) for each neighboring node may also be included if available, or otherwise be set to one. To reduce the burden of feedback, the link quality factors should

[1] This paper is prepared for presentation at IEEE SPS Workshop on DSP, Taos Ski Valley, NM, 1-4 August 2004. Part of this work was presented at ARL CTA Review meeting, Univ. of Delaware, March 2004. This work was supported in part by the ARL CTA program and NSF.

[2] The circuit and system design to achieve the symbol-level synchronization is currently under investigation by a joint team from UCR and UCLA. Symbol synchronization for multiple mobile relays is feasible provided that the symbol interval is relatively large. For a multicarrier system, each symbol may span 20 μs during which a radio signal propagates 6000 *m*. Hence, the time delays caused by the spatial distribution of mobiles are negligible provided that the mobiles are relatively clustered together. The accuracy of modern clock is well within 100 *ns*.

generally be dependent on large scale fading but not on small scale fading. Such a table can be established and maintained using the existing techniques for route discovery and maintenance [1-2]. The information stored in the route table does not allow any single node to independently determine a route from a source to a destination[3]. But this table is sufficient for the network to relatively easily find a route when such a route is needed. Generally, the more information is stored in route table, the faster a route can be found when needed, but the slower and the more expensive is the maintenance of route table. Our routing protocol is based on the next-hop information available in the route table at each node.

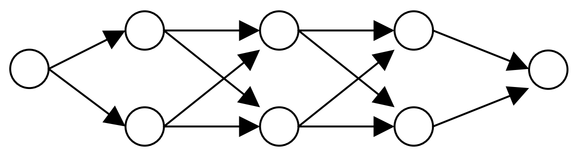

Figure 1: A schematic diagram of mobile parallel relays. Shown here is a route of four links (tiers) from source to destination, where each tier has N parallel relays. The relays in each tier are close to each other, and can reach by radio the relays in the adjacent tiers.

To establish a route of parallel relays such as the one shown in Figure 1, we start with the source node (the left most node). The source node must look for the relays in tier 1 (the nodes in the second left column). To achieve that, the source uses the information in its current route table and sends[4] out its intention, the source ID and the destination ID to those neighboring nodes that have a link to the destination. These neighboring nodes reply in the order specified by the source, and provide the pertinent part of their route tables to the source. The source then identifies *the nodes in tier 1* that have *the common next-hop nodes in tier 2*. These nodes in tier 1 are identified as the relays in tier 1. The source then informs the relays in tier 1 of the common next-hop nodes in tier 2. These relays are assigned with different sequence numbers by the source (according to the link quality factors between tier 2 and tier 3, which is available to the source from the information collected from the nodes in tier 2). A maximum number of relays in each tier should be set in the network. The first relay (with the best link quality) serves as the head relay.

To continue, the relays in tier 1 must then look for the relays in tier 2. The head relay in tier 1 sends out its intention, its ID and the destination ID to the common next-hop nodes in tier 2. The common next-hop nodes in tier 2 reply in the order specified by the head relay in tier 1, and supply the pertinent part of their route tables. The head relay in tier 1 works out *the nodes in tier 2* that have *the common next-hop nodes in tier 3*. These nodes in tier 2 are identified as the relays in tier 2. The head relay in tier 1 then informs the relays in tier 2 of the common next-hop nodes in tier 3. The relays in tier 2 are also given different sequence numbers by the head relay in tier 1.

The above process continues until the relays in the second last tier are identified and informed by its previous tier. To provide a further illustration, the nodes in each tier can be classified as follows.

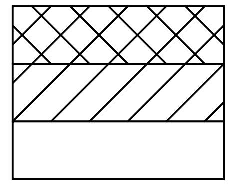

Figure 2: Classification of nodes at each tier.

Referring to Figure 2, the three rectangular blocks contain all the nodes in tier *t* that are neighboring the relays in tier *t*-1. The single and double shaded blocks contain the nodes in tier *t* that have at least one *next-hop node* to the destination. The double shaded block only contains the nodes that have the *common next-hop nodes* to the destination.

As shown in [4-7] and further in this paper, a route of parallel relays is robust against small scale fading caused by motions over a large fraction of wavelength, such a route needs not to be repaired as frequently as a route of single relays. The above dialogue between adjacent tiers only involves control information, and is hence much shorter in time than the data transmission. In fact, the data transmission from the source can start as soon as the head relay in tier 2 begins its dialogue with the nodes in tier 3. We assume here that the control channel is separated[5] from the data channel, which is an effective choice and is also practiced in the current cellular systems. If the control channel is the same as the data channel, the data transmission from the source needs to wait at least until the head relay in tier 2 completes its control dialogue (so that the relays in tier 1 is free from radio interference from tier 2).

The above algorithm establishes the feasibility of routing for parallel relays in a fairly arbitrary setting. In practice, such an algorithm may not be necessary. Each set of parallel relays may be organized by other more efficient means, e.g., locally within each cluster of nodes.

3. TRANSMITTING DATA THROUGH PARALLEL RELAYS

Once a route of parallel relays is established at tier 3 or beyond, data transmission may begin at the source. Upon notification from tier 1, the source sends out a packet of data to the relays in tier 1. The packet contains the data and

[3] Unless the destination is a next-hop node of the source.

[4] A brief and cooperative handshake (e.g., carrier sense multiple access) may first take place at the link layer to minimize radio collision.

[5] In frequency, time or spread-spectrum code.

various control information (such as relay ID, source ID, destination ID, packet ID, etc). The packet is also encoded for error correction and error detection. If any of the relays in tier 1 successfully[6] decodes the packet, each of the successful relays may use a space-time encoder[7] as specified by the relay sequence number [4-6]. The encoded packets from all the successful relays are simultaneously[8] transmitted to the relays in tier 2. Each of the relays in tier 2 does space-time decoding of the received packet. The successful relays in tier 2 forward the data to tier 3.

During the transmission of a packet at each link, it is possible that none of the receiving relays receives the packet correctly. In this case, retransmission is necessary. Let us have another look at the link between the source and tier 1. When the source hears the forwarded packet from tier 1, the source decides that the packet previously transmitted to tier 1 is correctly received by some of the relays in tier 1. If the source hears no signal[9] after a given period of time, the source retransmits the data packet. The process repeats until the packet from the source is received by at least one relay in tier 1. The same process of retransmission is applied to the link between tier 1 and tier 2 and all other subsequent links.

To transmit the next packet, the source needs to hear a control signal originated from tier 2. When the head relay in tier 2 hears the forwarded packet from tier 3, it sends out a "clear" control signal to the head relay in tier 1. Then, the head relay in tier 1 sends out "ready to receive" control signal to the source. Then, the source begins to send the next packet.

In general, after the relays in tier t receive the packets from tier t-1, the relays in tier t must wait for a "ready to receive" control signal from the head relay in tier t+1 unless such a control signal was received earlier. It is possible that many packets may have to be buffered at a tier due to a long delay of its downlink. When the buffer exceeds a limit, the tier simply stops sending "ready to receive". We have assumed that the nodes participating as parallel relays are homogenous and the synchronization at symbol level is achievable.

[6] Packet error detection is considered to be very reliable. In other words, the probability of undetectable errors can be easily made much smaller than the probability of error [3].

[7] Another simple choice is to use a distinct spread-spectrum code at each relay, which of course consumes an additional spectrum that could be used for other purposes.

[8] An alternative is to use the relays in each tier sequentially as needed pending on the previous transmission. This approach causes additional delay (especially when the link quality factors known to the relays do not reflect the actual link quality during each packet period).

[9] If the relays in each tier have a special local communication capacity with each other, the relays may be able to actively inform (with a minimum burden on the network) their previous tier if none of the relays decodes the packet successfully.

4. ANALYSIS OF NETWORK THROUGHPUT AND POWER SAVING

To analyze the network throughput and power saving, we assume the following:

a) There are N parallel relays in each tier except at the source and the destination.
b) The channel fading between a transmitting node and a receiving node of each link is statistically independent and identically distributed (i.i.d.) across all transmit-and-receive pairs within the link, and all links in the network have the same statistical property.
c) The channel fading factor is constant during the transmission of each single data packet. But the fading factor may change from packet to packet. (Multiple packets may be transmitted over multiple carriers of different fading factors.)

The bit error rate (BER) at the ith receiving node in response to k transmitting nodes is denoted by $p_i(k)$. With a full-diversity space-time modulation method used at the transmitting nodes, the averaged BER over Rayleigh fading channels for a class of symbol modulation methods is known to be proportional to $1/SNR^k$ for large (averaged) SNR. If a block code and a hard decision decoding are used, the packet loss rate (PLR) is

$$P_i(k) = \sum_{j=c+1}^{J} \binom{J}{j} p_i^{\,j}(k)\left(1-p_i(k)\right)^{J-j} \qquad (1)$$

where c is the number of correctable error bits. Following the Appendix, we can show that for large SNR, the averaged PLR is also proportional to $1/SNR^k$, i.e.,

$$P_k \mathrel{\hat{=}} E\left(P_i(k)\right) \propto \frac{1}{SNR^k}$$

where E denotes expectation.

Assuming that the source has only one transmitter, the average probability that g receiving relays in tier 1 lose a packet is

$$G_1(g) = \binom{N}{g} P_1^g \left(1-P_1\right)^{N-g} \qquad (2).$$

More generally, for the relays in tier t and $t \geq 2$, the average probability that g receiving relays lose a packet can be shown to be

$$G_t(g) = \binom{N}{g} \sum_{k=1}^{N} P_k^g \left(1-P_k\right)^{N-g} \frac{G_{t-1}(N-k)}{1-G_{t-1}(N)} \qquad (3)$$

where the denominator $1-G_{t-1}(N)$ is due to the fact that the relays in tier t-1 do not forward a packet until at least one of them received the packet correctly.

It is clear that $G_t(N)$ measures the average probability of link failure in tier t. Following the Appendix, we can show that for large SNR,

$$G_t(N) \propto 1/SNR^{d(t)}$$

where

$$d(1) = N$$
$$d(2) = 2N - 1$$
$$d(t) = \min\left(tN - t + 1, \quad N^2\right) \text{ for } t \geq 3 .$$

For $t \geq N+1$, $d(t) = N^2$, i.e., the diversity gain is N^2. Note that with only two relays at each link, we have that for $t \geq 3$, $G_t(2) \propto 1/SNR^4$.

Consider the use of Golay (24,12) code for error correction and detection [3]. The following figure shows the packet loss rate (PLR) as a function of *SNR* where two relays are used at each tier. The Alamouti code is used for space-time modulation. QPSK symbol modulation is assumed. The top curve in Figure 2 is the PLR at tier 1, with a slope equal to 2. The second curve is the PLR at tier 2, with a slope equal to 3. The third curve is the PLR at tier 3, with a slope equal to 4. All other curves have the slope equal to 4 as predicted by theory.

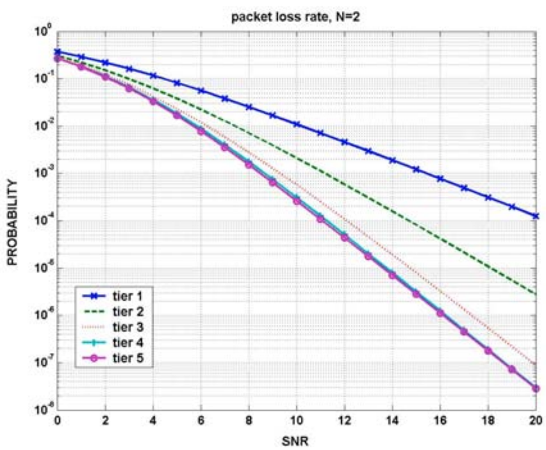

Figure 2: Packet loss rates at different tiers where two relays are used at each tier.

The average delay at each link can be shown to be

$$T_t(N) = \frac{\Delta}{1 - G_t(N)}$$

where Δ is the time of each transmission of a packet, successfully or not. The network throughput can be measured by a normalized inverse of $T_t(N)$, i.e.,

$$C_t(N) = 1 - G_t(N) .$$

Assuming $P_{k+1} < P_k$, the gain of the network throughput from serial relaying to parallel relaying is

$$K_t(N) \triangleq \frac{C_t(N)}{C_t(1)} = \frac{1 - G_t(N)}{1 - P_1}$$
$$> \frac{1 - P_1^N}{1 - P_1} = 1 + P_1 + P_1^2 + \cdots + P_1^{N-1}$$

which is significant when P_1 is not very small. Figure 3 illustrates the lower bound of the throughput gain.

To consider the power saving from serial relaying (N=1) to parallel relaying (N>1), we let the total transmission power to be upper bounded by a constant independent of *N*. The orthogonal space-time modulation is used in all cases. Figure 4 shows the PLR at tier 5 for different values of *N* at each tier. We see a significant power saving by using $N > 1$. To keep the PLR at 1% from N=1 to N=2, there is almost a 10 *dB* reduction of *SNR*. It is important to note that from N=1 to N=2, there is no additional cost of bandwidth. Beyond N=2, there is a penalty factor of bandwidth (between one and two) for using orthogonal space-time modulation [5], [9]. The power saving becomes insignificant when N >4.

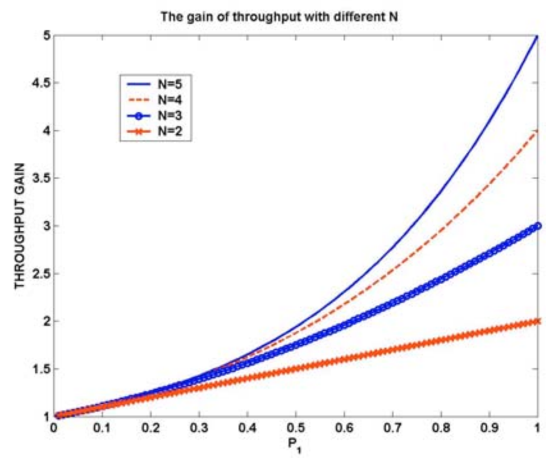

Figure 3: The lower bound of throughout gain from serial relaying to parallel relaying as function of the packet loss rate between two single nodes.

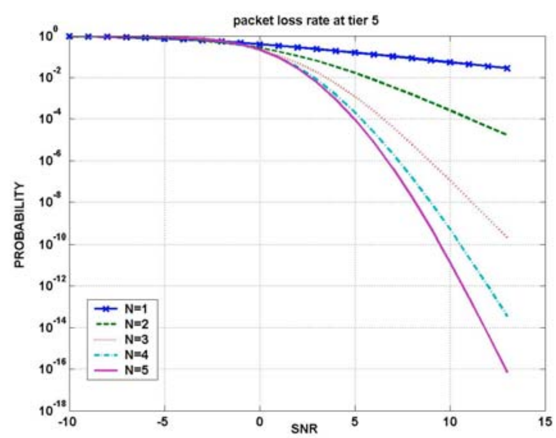

Figure 4: Packet loss rates at tier 5 for different numbers (N) of parallel relays at each tier. The total transmission power at each tier is kept to be independent of N.

5. CONCLUSION

Our study suggests that mobile parallel relays are feasible from the networking point of view. The potential benefits of mobile parallel relays are significant. With a reduced burden on networking traffic, mobile parallel relays may yield more than 10 *dB* power saving during data transmission in a highly mobile environment.

6. APPENDIX

In this appendix, we will prove that for any finite *n*, $E\left(p_i^n(k)\right) \propto \rho^k$ where $\rho = 1/SNR$. Then, it follows from (1) that $E\left(P_i(k)\right) \propto \rho^k$. Furthermore, it follows from (2) and (3) that

$$G_1(l) \propto \rho^l, \; l=0,1,2,\ldots,N$$

$$G_2(l) \propto \begin{cases} \rho^{l+N-1} & l>0 \\ 1 & l=0 \end{cases}$$

$$G_3(l) \propto \begin{cases} \rho^{l+2(N-1)} + \rho^{lN} & l>0 \\ 1 & l=0 \end{cases}$$

$$G_t(l) \propto \begin{cases} \rho^{l+(t-1)(N-1)} + \rho^{lN} & l>0 \\ 1 & l=0 \end{cases}.$$

In particular, we have

$$G_t(N) \propto \rho^{N+(t-1)(N-1)} + \rho^{N^2}.$$

Proof of $E\left(p_i^n(k)\right) \propto \rho^k$: For coherent M-PSK with Gray coding, the (conditional) bit error rate (BER), $p_i(k)$, can be approximated by the symbol-error rate (SER) $\frac{1}{\log_2 M} p_s(k)$ [10]. From [11], the SER is given by

$$p_s(SNR_{in}) = \frac{1}{\pi}\int_0^{\pi-\pi/M} \exp\left(-\frac{SNR_{in}\sin^2(\pi/M)}{\sin^2\theta}\right) d\theta$$

where $SNR_{in} = \chi SNR$ is the instantaneous signal-to-noise ratio, and χ is a random variable of χ^2_{2k} distribution. As shown in [11], we have

$$L\left(SNR_{in}\right) < p_s(SNR_{in}) < U\left(SNR_{in}\right)$$

where

$$L(SNR_{in}) = \sum_{j=2}^{T_1} a_j \exp(-b_{j-1}SNR_{in}) + \sum_{j=1+T_1}^{T} a_j \exp(-b_j SNR_{in})$$

$$U(SNR_{in}) = \sum_{j=1}^{T_1} a_j \exp(-b_j SNR_{in}) + \sum_{j=1+T_1}^{T} a_j \exp(-b_{j-1} SNR_{in})$$

with $a_i = \frac{\theta_i - \theta_{i-1}}{\pi}$, $b_i = -\frac{\sin^2(\pi/M)}{\sin^2\theta_i}$, and

$$0 = \theta_0 < \theta_1 <,\ldots,< \theta_{T_1} = \pi/2 <,\ldots,< \theta_T = \pi - \pi/M.$$

The value of *T* affects the tightness of the bounds. It then follows that

$$E\left[L^n(SNR_{in})\right] \le E\left[p_s^n(SNR_{in})\right] \le E\left[U^n(SNR_{in})\right]$$

where the expectation is carried over the distribution of χ, and both the lower and upper bounds have the following common structure:

$$\sum \alpha_i E\left[\exp(-\beta_i \chi SNR)\right].$$

The PDF of χ is $\frac{1}{\Gamma(2k/2)2^{2k/2}} \chi^{2k/2-1} \exp(-\chi/2)$. It follows that $E\left[\exp(-\beta_i SNR\chi)\right] = \left(1+2\beta_i SNR\right)^{-k}$ and hence

$$E\left[U^n(SNR_i)\right] = \sum \frac{\alpha_{i,U}}{\left(1+2\beta_{i,u} SNR\right)^k}$$

$$E\left[L^n(SNR_i)\right] = \sum \frac{\alpha_{i,L}}{\left(1+2\beta_{i,L} SNR\right)^k}.$$

With high SNR, both the upper and lower bounds become proportional to $1/SNR^k = \rho^k$ and hence $E\left(p_i^n(k)\right) \propto \rho^k$.

7. REFERENCES

[1] C. E. Perkins, ed., *Ad hoc networking*, Addison-Wesley, 2001.

[2] C. – K Toh, *Ad hoc mobile wireless networks – protocols and systems*, Prentice-Hall, 2002.

[3] W. W. Peterson, and E. J. Weldon, Jr., *Error-Correcting Codes*, 2nd edition, The MIT Press, 1972.

[4] Y. Hua, Y. Mei, Y. Chang, "Parallel wireless mobile relays with space-time modulation", *IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, Sept. 28- Oct 1, 2003.

[5] Y. Hua, Y. Mei, Y. Chang, "Wireless antennas – making wireless communications perform like wireline communications," *IEEE Topical Conference on Wireless Communication Technology*, pp. 1-27, Honolulu, Hawaii, Oct 15-17, 2003.

[6] Y. Chang and Y. Hua, "Application of space-time linear block codes to parallel wireless relays in mobile ad hoc networks," *Proc of Asilomar Annual Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 10-12, 2003.

[7] P. A. Anghel, G. Leus, and M. Kaveh, "Multi-user space-time coding in a cooperative networks", *Proc of IEEE ICASSP'2003*, Hong Kong, May 2003.

[8] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, submitted in Jan 2002.

[9] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge, 2003.

[10] J. G. Proakis, *Digital Communications*, 4th edition, McGraw-Hill, 2001.

[11] M. Chiani, D. Dardari and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Transactions on Wireless Communications*, pp. 840-845, No. 4, Vol. 2, July 2003.