High Throughput Route Selection in Multi-Rate Ad Hoc Wireless Networks

Baruch Awerbuch, David Holmer, Herbert Rubens
{baruch, dholmer, herb}@cs.jhu.edu

***Abstract*— An ad hoc wireless network is an autonomous self-organizing system of mobile nodes connected by wireless links where nodes not in direct range communicate via intermediate nodes. Modern wireless devices, such as those that implement the 802.11b standard, utilize multiple transmission rates in order to accommodate a wide range of channel conditions. We provide a general theoretical model of the attainable throughput in multi-rate ad hoc wireless networks. This model is derived from a detailed analysis of the physical and medium access control layers.**

The traditional technique used by most existing ad hoc routing protocols is to select minimum hop paths. These paths tend to contain long range links that have low effective throughput and reduced reliability. We present the *Medium Time Metric* (MTM) that selects optimal throughput paths and tends to avoid long unreliable links. Our NS2 simulation environment consists of a high mobility 802.11b network with many simultaneous TCP connections. These simulations show that our MTM yields an average total network throughput increase of 20% to 60%, depending on network density, over the traditional hop count metric. By combining the MTM with the Opportunistic Auto Rate (OAR) protocol, an increase of 100% to 200% is obtained over the traditional route and rate selection techniques. This work stresses the importance of inter-layer communication in wireless networks. Our results show that link rate information from the medium access control layer can be utilized by routing protocols to significantly increase network performance.

***Index Terms*—multi-rate, ad hoc, wireless, routing, routing metric, 802.11b, cross layer interaction.**

I. Introduction

AD HOC wireless networks are self-organizing multi-hop wireless networks where all nodes take part in the process of forwarding packets. One of the current trends in wireless communication is to enable devices to operate using many different transmission rates. Many current and proposed wireless networking standards have this multi-rate capability. These include the 802.11b [1], 802.11a [2], 802.11g draft, and HiperLAN2 [3] standards. The reason for this multi-rate capability stems directly from some of the fundamental properties of wireless communication.

Due to the physical properties of communication channels, there is a direct relationship between the rate of communication and the quality of the channel required to support that communication reliably. Since distance is one of the primary factors that determines wireless channel quality, there is an inherent trade-off between high transmission rate and effective transmission range.

Authors are with the Computer Science Department, Johns Hopkins University, Baltimore, Maryland. Technical Report Version 2 - 3/12/2003

This range speed trade-off is what has driven the addition of multi-rate capability to wireless devices. Consumer demands for wireless devices always include both higher speed and longer range. Unfortunately a single rate represents a single trade-off point between these two conflicting goals. Since multi-rate devices support several rates, they provide a wide variety of trade-offs available for use. This gives them a great deal of flexibility to meet the demands of consumers. This added flexibility is the primary driving force behind the adoption of multi-rate capability. It is also reasonable to assume that this type of capability will also be present in future wireless networking standards.

While multi-rate devices provide increased flexibility, they cannot change the inherent trade-off between speed and range. Both high speed and long range cannot be achieved simultaneously. Long range communication still must occur at low rates, and high-rate communication must occur at short range. This multi-rate capability merely provides a number of different trade-off points. Multi-rate devices must have protocols that select the appropriate rate for a given situation.

In infrastructure based networks, all communication takes place between nodes and access points. In this case, an additional protocol required to support multi-rate is necessary only at the medium access control (MAC) layer. Single rate nodes already have the ability to select the best access point based on the received signal strength. Thus the only additional task necessary is that of selecting the actual rate used to communicate. Since the distance between the user and the access point is dictated by the physical geometry of the network, the rate selection task must react to the existing channel conditions. In other words, the only option available to a wireless device is to select the fastest modulation scheme that works reliably.

However, this is no longer the case in ad hoc multi-hop wireless networks. In these networks, the routing protocol must select from the set of available links to form the path between the source and the destination. While in single-rate networks all links are equivalent, in multi-rate networks each available link may operate at a different rate. Thus the routing protocol is presented with a much more complex problem. Which set of trade-offs does it choose? Long distance links can cover the distance to the destination in few hops, but then the links would be forced to operate at a low speed. Short links can operate at high rates, but more hops are required to reach the destination. In addition, the path selected by the routing protocol will not only affect the packets moving along that path, but will affect the level of congestion at every node within the interference range of the path as well.

Our Contribution. We provide a general theoretical model of the attainable throughput in multi-rate ad hoc wireless networks. This model is derived from a detailed analysis of the physical and medium access control layers. The traditional technique used by most existing ad hoc routing protocols is to select minimum hop paths. These paths tend to contain long range links that have low effective throughput and reduced reliability. We present the *Medium Time Metric* (MTM) that selects optimal throughput paths and tends to avoid long unreliable links. The MTM minimizes the total medium time consumed sending packets from a source to a destination. This results in an increase in total network throughput.

The rest of the paper is organized as follows. Section II summarizes related work. We further define our network model and assumptions in Section III. In order to fully understand the effects of the physical and MAC layers on network throughput, we present a detailed analysis in Section IV. We examine shortcomings of a few common route selection techniques in Section V. In Section VI we present a theoretical model of throughput in multi-rate networks and derive an optimal route selection heuristic. We present our Medium Time Metric in section VII. In Section VIII we discuss the potential complexities involved in incorporating the MTM into existing protocols, and the modifications made to the DSDV [4] routing protocol used in our simulations. The results of these simulations are provided in Section IX and we conclude in Section X.

II. RELATED WORK

A. Ad Hoc Routing Protocols

A large number of routing protocols have been proposed by the ad hoc wireless networking community. Typically these have adopted one of two major strategies: on-demand such as in AODV [5] and DSR [6], and proactive such as in DSDV [4] and OLSR [7]. The vast majority of these protocols have used a shortest path algorithm with a hop count metric (min hop) to select paths. While min hop is an excellent criteria in single-rate networks where all links are equivalent, it ignores the trade-offs present in multi-rate networks. It should be possible to enhance the multi-rate network performance of almost any existing shortest path based protocol by adapting it to use our medium time metric.

B. Signal Stability Based Ad Hoc Routing Protocols

In [8] the authors show that the minimum hop path generally contains links which exhibit low reliability. In [9] and [10] the authors present routing protocols which are based on signal stability rather then just shortest path in order to provide increased path reliability. In this work signal stability information is used not only to increase path reliability, but also to increase network throughput.

C. MAC Layer

Since our proposed solution is derived from properties of the MAC and physical layers, it is important to understand existing MAC layer techniques. The IEEE 802.11 standard [11] defines the most commonly used MAC protocol in ad hoc wireless networks. 802.11 based devices are used because of their widespread availability, low cost, and 802.11's ability to provide distributed medium access control when operated in "ad hoc" mode. This mode causes the stations to use the Distributed Coordination Function (DCF) protocol that operates using Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The MAC protocol operates using an exchange of control frames for each data packet. This exchange consists of a Request To Send (RTS), Clear To Send (CTS), Data Packet (DATA), and Acknowledgement (ACK) frame.

The method of rate selection in multi-rate capable networks has been left unspecified by the 802.11 standards. As a result, several auto rate protocols have been proposed. The most commonly used protocol is Auto Rate Fallback (ARF). ARF was originally developed for Lucent's WaveLAN II devices [12], and was later enhanced for 802.11b devices [13]. ARF operates using the link level ACK frames specified by the 802.11 standard. Each node increases the rate it is using to communicate with its neighbor after a number of consecutively received acks, and decreases the rate after a number of consecutively missed acks. The advantage of this technique is that it is easy to implement because it is purely sender based, requires no modifications to the 802.11 standard.

As an alternative, the Receiver Based Auto Rate (RBAR) protocol was presented in [14]. RBAR allows the receiving node to select the rate. This is accomplished by using the SNR of the RTS packet to choose the most appropriate rate and communicating that rate to the sender using the CTS packet. This allows much faster adaptation to the changing channel conditions than ARF, but requires some modifications to the 802.11 standard.

The Opportunistic Auto Rate (OAR) protocol, which is presented in [15], operates using the same receiver based approach, but allows high-rate multi-packet bursts to take advantage of the coherence times of good channel conditions. These bursts also dramatically reduce the overhead at high rates by amortizing the cost of the contention period and RTS CTS frames over several packets. By picking appropriate sized bursts, OAR also changes the fairness characteristic from each node sending an equal number of packets to each node getting an equal allocation of medium time. This produces a dramatic increase in overall throughput when links of multiple rates operate together in the same space. OAR also requires modifications to the 802.11 standard.

III. NETWORK MODEL

A. Network Assumptions

This work relies on a few specific network assumptions. We assume that the ISO/OSI physical layer is capable of operating using multiple rates. We also assume that the ISO/OSI MAC layer is capable of selecting the rate used by the physical layer. In addition, we assume that the MAC layer is capable of providing information to the ISO/OSI network layer that indicates the selected rate. The network layer can then use this information to improve its routing decisions. This work stresses the importance of inter-layer communication in wireless networks. We demonstrate that information from lower layers can be utilized to enhance overall performance.

B. Multi-Rate Model

The multi-rate model presented in this paper is based on the 802.11b standard [1]. The topics discussed here apply to other multi-rate standards, but all examples, ranges, and rates shown in this work are based on 802.11b.

Throughout the remainder of the paper we present the results of a number of NS2 [16] simulations. In order to simulate multi-rate 802.11b, we started with the ns-2.1b7a code base and the multi-rate extensions available from the Rice Networks Group [17] that contain implementations of the RBAR and OAR protocols. The 802.11 MAC and physical wireless parameters were further modified to match the published specifications of a Lucent ORiNOCO PC Card [18], a commonly used 802.11b wireless adapter (see Table I). Since the carrier sense (CS) threshold specification is not published, we provide an estimate. This estimate was produced by setting the difference between the carrier sense threshold estimate and the 1.0 Mbps receive threshold equal to the difference between the NS2 default carrier sense threshold (-78 dBm) and default receive threshold (-64 dBm).

Figure 1 and Table II show the ranges resulting from these simulation parameters. Real world ranges are considerably smaller due to non-zero system loss, additional noise sources, obstructions, and propagation effects beyond the simple two ray ground model. The results presented here should be valid for any set of ranges with similar proportions regardless of magnitude.

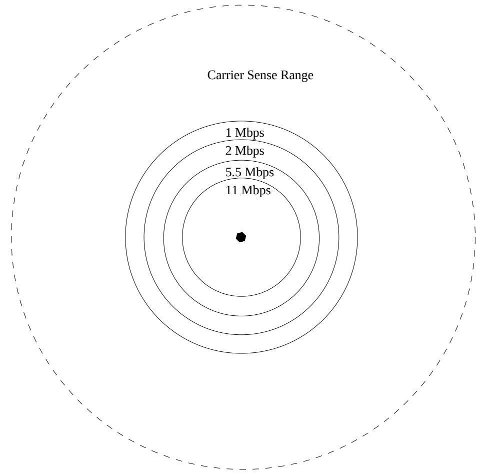

Fig. 1. 802.11b Ranges

IV. Throughput Phenomena in Multi-Rate Ad Hoc Wireless Networks

The total network throughput attainable in multi-rate ad hoc wireless networks is a result of the combined behavior of the medium access control protocol, routing protocol, and physical properties of a wireless network. In order to provide an understanding of how this combined behavior affects network throughput, we examine several different phenomena.

A. Non-Overlapping Transmissions

Ad hoc wireless networks by nature use a broadcast medium. This means that any transmission made by a node simultaneously propagates to all other nodes in range. The downside of this property is that even if a node is sending packets to only one of its neighbors, those packets affect every other node in range. Furthermore, if two nodes transmit simultaneously, both transmissions will overlap and become garbled on the medium causing a receiver to be unable to successfully receive either packet. As a result, only a single transmission can occur at a time within range of the intended receiver.

TABLE I

NS2 Simulation Parameters

Parameter	Value
Frequency	2.4 GHz
Transmit Power	15 dBm
11.0 Mbps Receive Threshold	-82 dBm
5.5 Mbps Receive Threshold	-87 dBm
2.0 Mbps Receive Threshold	-91 dBm
1.0 Mbps Receive Threshold	-94 dBm
Carrier Sense Threshold	-108 dBm
Capture Threshold	10
Propagation Model	Two Ray Ground
System Loss	0 dBm

B. Transmission Deferral

The MAC protocol is responsible for providing channel access arbitration and ensuring that nodes defer sending to avoid interfering with a transmission in progress. The 802.11 MAC protocol uses two mechanisms for deferral. The first mechanism used is carrier sensing, which means that the node listens to the medium in order to detect when another transmission is in progress. If it hears a transmission it defers until the medium is idle. Only nodes that are within carrier sense range of a sender will be able to successfully use this method to avoid collisions. The second mechanism is referred to as virtual carrier sense, and it is provided by a control frame exchange. A Request To Send (RTS) control frame is transmitted by the sender when it has a data packet to deliver. If the receiver is not already deferring, it responds with a Clear To Send (CTS) control frame. Any node that overhears an RTS or CTS is notified of the packet transmission, and will then defer for the duration of the transmission. This additional mechanism is particularly useful in cases where nodes near the receiver cannot carrier sense the transmission because of obstacles or other propagation effects.

TABLE II

802.11b Ranges

Rate (Mbps)	Maximum Range
11.0	399 m
5.5	531 m
2.0	669 m
1.0	796 m
CS	1783 m

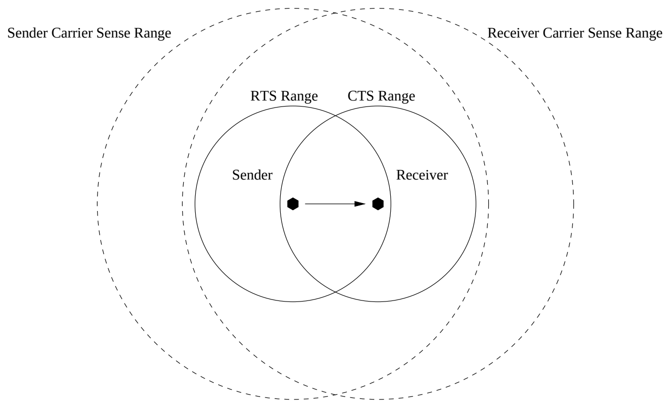

Fig. 2. 802.11 Range

Figure 2 illustrates the ranges of these two mechanisms according to the specified communication model.

In addition to providing medium reservation, the RTS and CTS frames also serve other purposes. The first is fast collision resolution, which is necessary because of the lack of collision detection hardware in wireless devices. The second is that the RBAR and OAR rate selection protocols use the RTS frame to provide a direct measurement of the current channel quality. The receiver can then select the most appropriate rate and notify the sender using the CTS frame. Since the receiver is able to select the rate every time it receives an RTS frame, it is able to respond quickly to variations in channel conditions.

C. Medium Access Control Overhead

The MAC protocol is responsible for providing channel access, which incurs a significant amount of overhead. In 802.11 this overhead is composed of three primary components: time spent transmitting control frames, random back-off time spent during contention resolution, and time wasted as a result of collisions.

Collision detection, which is used in Ethernet [19] networks is impossible in wireless networks. In an effort to reduce the total overhead, 802.11 spends a significant amount of medium time sending control frames that are designed to help avoid costly data packet collisions. As a result, medium access control is more expensive in the wireless environment than in the wired environment.

The result of this MAC overhead is that the effective throughput is less than the link rate. Table III shows the results of a simple NS2 experiment where 1472 byte UDP packets were flooded across a single link. The time spent for data and overhead in 802.11b are shown in Figure 3. The 802.11 MAC overhead is significant, particularly for the higher rate links. The effective throughput of an 11 Mbps link is less than half the link rate. Only the contents of the DATA and ACK frames are transmitted at the selected link rate, the rest of the exchange occurs at the 1 Mbps base rate. As a result, the MAC overhead is almost constant per packet. Therefore, the effective link rate is determined by the amount of time spent transmitting the data contents of each packet. We see a greater reduction in effective throughput for faster links because the time necessary to send a packet is inversely proportional to the rate of link. In other words, the data transmission time is small for fast links, the proportion of time consumed by the fixed overhead is large.

TABLE III
SINGLE LINK THROUGHPUT

Rate (Mbps)	Throughput (Mbps)
11.0	4.55
5.5	3.17
2.0	1.54
1.0	0.85

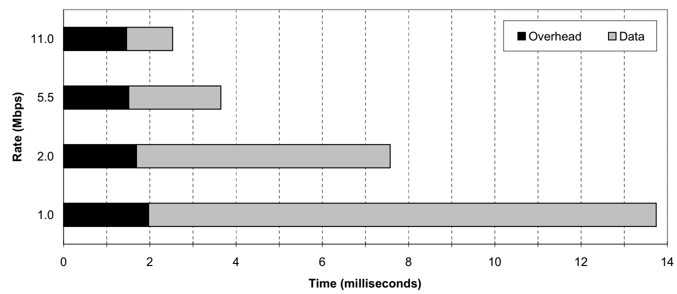

Fig. 3. Medium Times for 802.11b Transmissions

D. Simultaneous Transmissions

When considering the total throughput in the wireless network, it is important to consider the number of non-interfering transmissions that can simultaneously exist as well as the rate at which each transmission is occurring. Unfortunately, the number of simultaneous transmissions is determined by the physical network topology and the transmission power level. The greater the geographic size of the network the greater the number of possible simultaneous transmissions. A protocol cannot control the physical configuration of nodes in the network, but it can control the rate at which the nodes transmit data.

Given a network where three simultaneous transmissions can occur, if these transmissions are sent at 1 Mbps, which is the lowest 802.11b transmission rate, a maximum of 3 Mbps of total network throughput could be obtained. Consider the same network, but with transmissions occurring at 11 Mbps. This would result in a total network throughput of 33 Mbps, which is significantly greater.

E. Link Rate vs. Hops Trade-Off

One approach to increasing throughput would be to configure all the nodes in the network to operate only at the highest transmission speed. This would ensure that the network would always operate at the maximum combined simultaneous rate. This approach may run into problems because of the inherent trade-off between the transmission rate and effective transmission range (see Figure 1).

In multi-hop ad hoc networks, packets must frequently traverse several hops to travel from the source to the destination. By using slow links that have high effective range, the distance between the source and destination can be covered using a small

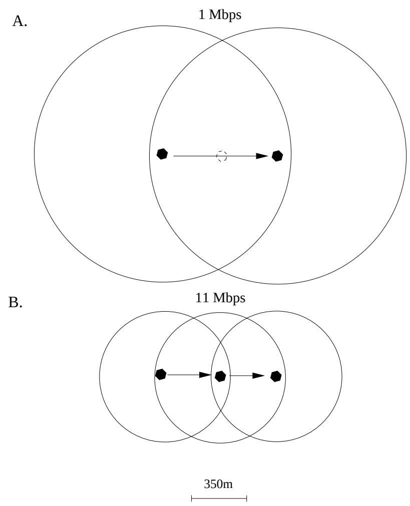

Fig. 4. Path Selection Options

number of hops. If we avoid using all but the fastest links, we reduce the effective range of every node. One major drawback of this approach is that we run the risk of disconnecting components of the network. Even if we do not disconnect the network, we increase the number of hops required to cover the distance from the source to the destination.

F. Hops vs. Throughput Trade-Off

Consider the following example where the source and destination are barely within range of one another (see Figure 4-A). In this configuration the source can reach the destination in one transmission at the lowest rate. A single link is by definition the minimum hop path between the source and the destination since no other path can be shorter. While sending the packet directly to the destination would result in the least number of transmissions, the transmission would occur at the slowest possible speed, requiring all of the other nodes in this neighborhood to defer transmitting for the longest possible time. As we previously discussed, transmitting at this rate will limit the overall throughput attainable in the network.

Now consider the same situation except an additional node is located between the source and the destination (see Figure 4-B). The source and destination can still communicate directly through one low speed transmission, but now an additional option exists. The traditional minimum hop path algorithm would not consider this configuration any differently from the previous, since routing through the intermediary node would only increase the hop count. The speed of each of the two transmissions would be 11 Mbps as opposed to the single 1 Mbps transmission selected by the minimum hop approach. This would provide an effective bandwidth along the path of 2.38 Mbps by utilizing two 11 Mbps hops as opposed to 0.85 Mbps across the single 1 Mbps link. This represents almost a three fold increase in throughput (see Tables III and IV).

The previous example suggests that choosing routes that use high-rate links is strictly better then those that use low-rate links. While this is true in many individual situations (including the one above), there are other factors to consider. In the previous example, two 11 Mbps links were used to provide increased

TABLE IV
TWO HOP PATH THROUGHPUT

Link Rate (Mbps)		Path Throughput
1^{st} Hop	2^{nd} Hop	(Mbps)
11.0	11.0	2.38
11.0	5.5	1.86
11.0	2.0	1.15
5.5	5.5	1.59
5.5	2.0	1.04
2.0	2.0	0.77

throughput over the single 1 Mbps link. Despite the fact that all of the links in the path operate at 11 Mbps, the throughput of the path is only a fraction of 11 Mbps. This is because only a single transmission can occur at a time in the same area. For the packets to traverse the two 11 Mbps hops, the source would have to alternate with the forwarding node. In other words, the nodes need to take turns transmitting. This coordination is handled by the medium access control layer.

In this simple example, the two 11 Mbps hops are strictly better than the single 1 Mbps hop, but this might not be the case if the choice is between ten 11 Mbps hops and a single 1 Mbps hop. There are several reasons why this is true. When packets are sent along a path in a multi-hop network, the adjacent transmissions are competing for access to the medium. By sending across many hops, the throughput along the path becomes a fraction of the capacity of the links. In Figure 5 nodes 1 and 8 are communicating along a path. The diagram shows the nodes that are affected by the transmission of node 4 while it is forwarding the packet on to node 5 along the path. In this example all eight nodes are being affected by the single transmission that is taking place and they all must defer from sending until the transmission completes.

In this example, nodes 2 through 6 are all in carrier sense range of node 4, which is transmitting. These nodes all defer until the transmission completes. Node 7 on the other hand is in carrier sense range of the receiver but not the sender. Node 7 can carrier sense the receiver's CTS packet, but will not be able to carrier sense the actual transmission. This will cause node 7

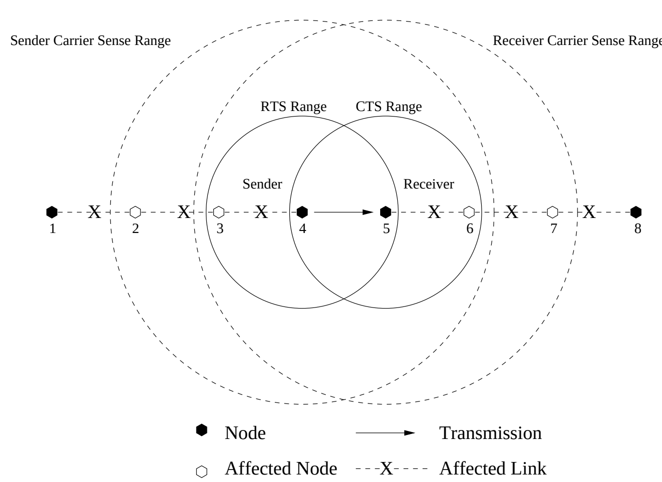

Fig. 5. Effect of Transmission on Other Nodes

to defer for an extended inter-frame spacing, which may not be long enough for the transmission from node 4 to 5 to complete. If node 7 begins transmitting it could potentially cause a collision. This example shows that the 802.11 MAC protocol has not solved the hidden terminal problem [20]. Another interesting aspect of this example is the effect of the transmission on nodes 1 and 8. Both of these nodes are out of the carrier sense ranges of both the sender and the receiver (nodes 4 and 5 respectively). As a result they appear to be unaffected by the transmission that is taking place. While it is true that these nodes could communicate with any other node outside of the current transmission neighborhood, in this particular example they are attempting to communicate along the path between node 1 and 8. Since nodes 2 and and 7 are currently deferring as a result of the transmission, any RTS initiated by nodes 1 or 8 would receive no reply. As a result nodes 1 and 8 will also need to defer until the transmission from node 4 to 5 completes. This example shows the broad impact that a single transmission has on nodes along the path as well as on other nodes in the immediate vicinity.

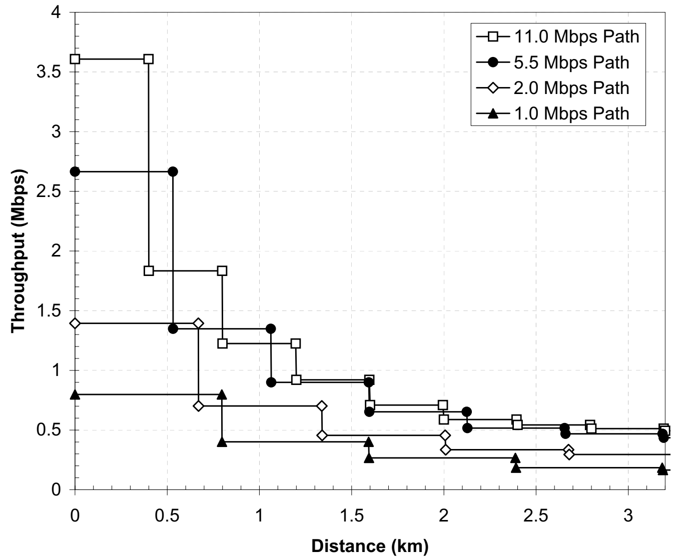

Fig. 6. Throughput Loss Along a Path

G. Quantitative Evaluation of Throughput Loss

An additional example shows a more quantitative evaluation of the throughput loss along a path. Figure 6 contains the results of an NS2 simulation that was conducted to explore the throughput loss of a single TCP connection along a path where each link operates at the same rate. Simulations were conducted for each of the four 802.11b link rates. The results show the throughput across the path vs. the distance (or length) of the path. As the length of the path increases the number of hops required to traverse the distance also increases. Since the throughput drops as the number of hops increases, the throughput drops in steps. The width of each step is equal to the effective transmission range at the given rate.

Since high-rate links have a shorter effective range, a greater number of hops is required to cover the same distance as a smaller number of lower rate hops. This is indicated in the graph since the high-rate throughput drops multiple times for each decrease in the low-rate throughput. There are a couple of interesting observations that are evident in this graph. The first observation is that the lines intersect. This means that at certain distances more throughput can be obtained using lower speed links then higher speed links. A specific example of this occurs at 0.4 km. Notice the throughput obtained by the 5.5 Mbps path is greater than that of the 11 Mbps path. This occurs because the 11 Mbps path needed to traverse 2 hops at this distance, while the 5.5 Mbps path still consists of a single hop. This shows that traversing high speed links does not always achieve the highest throughput in all cases. Another interesting observation is that after approximately 2.2 km the speeds seems to plateau. This is due to spatial reuse. As the path becomes longer, multiple transmissions can take place simultaneously along the path. This allows the throughput to reach a steady state, where additional distance does not cause any significant decrease in throughput. It is also important to notice that at this distance the throughput of the links increases as the link speed increases. This suggests that even though high link rate paths must traverse more links to reach the same distance, they still provide more throughput.

H. Temporal Fairness

In addition to low path throughput, there are other detrimental effects of sending packets at slow transmission speeds. The standard 802.11 MAC protocol attempts to provide fairness to individual senders on a per packet basis. This means that if there are two senders near each other and they are continuously trying to send packets, they should end up sending approximately the same number of packets. In multi-rate networks, there is no guarantee that these two senders are sending at the same rate. Since the MAC protocol is only attempting to be fair with regard to the number of transmissions, slow senders dominate the medium time. One technique for dealing with this problem involves redefining the MAC fairness model. Temporal fairness would provide an equal share of medium time between senders independently of their transmission rate. There has already been work which explores this option.

The Opportunistic Auto Rate (OAR) protocol provides temporal fairness with regard to medium time by allowing senders who send at a high-rate to send as many packets as required to equal the transmission time of a single packet at a low-rate. Basically, this results in every sender having an equal opportunity to transmit and for each sender to be able to transmit for the same amount of medium time. This is a dramatic improvement in efficiency over the existing 802.11 fairness model.

A simulation was run in the NS2 network simulator with nodes arranged as indicated in Figure 7. The simulation consisted of two nodes flooding packets to their destination. One sender was sending at 1 Mbps and the other was sending at 11 Mbps. All nodes in the simulation were within range of each other and were contending for access to the medium. The simulation was conducted with both the OAR and RBAR MAC protocols and the results shown in Table V are averaged over several random number seeds.

As seen in the results, the OAR provides almost two and a half times the total throughput of RBAR. This indicates that temporal fairness is extremely important for achieving high throughput in ad hoc networks. The RBAR results, which are

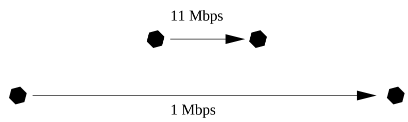

Fig. 7. Temporal Fairness Simulation Node Configuration

representative of the current 802.11 MAC, indicate that even if some of the routes in the network are operating at high link speeds, the total network throughput will still be low as a result of low speed links dominating network medium time. We conclude that in order to achieve high throughput, not only will the routing protocol need to be selecting high speed links, but the medium access control protocol will have to provide temporal fairness to ensure that low speed links do not gain an unfair share of the medium time.

V. Traditional Route Selection Techniques

In this section we discuss a few of the traditional route selection techniques used in both wired and wireless networks. We focus on the shortcomings of these techniques when they are applied to multi-rate wireless networks.

A. Minimum Hop Path

Most existing ad hoc routing protocols have utilized hop count as their route selection criteria. This approach minimizes the total number of transmissions required to send a packet on the selected path. This metric is appropriate in single-rate wireless networks because every transmission consumes the same amount of resources. However, in multi-rate networks this technique has a tendency to pick paths with both low reliability and low effective throughput.

1) Throughput Loss: In multi-rate wireless networks, the selection of minimum hop paths typically results in paths where the links operate at low rates. This is because the shortest path contains the fewest number of nodes between the source and destination. Fewer intermediate nodes corresponds to longer links in order to cover the same distance. Since distance is one of the primary factors that determines channel quality, the long links have low quality, and thus operate at low rates. So given the opportunity, in an effort to minimize the number of hops, shortest path selection protocols will pick paths composed of links close to their maximum range that must operate at the minimum rate.

Not only do the low link rates produce a low effective path throughput, but as a result of the shared wireless medium, this path selection degrades the performance of other flows in the network. This occurs due to the large amount of medium time required to transmit a packet at a slow link speed. All nodes within interference range of the transmission must defer while it takes place. Thus, slow transmissions reduce the overall network throughput by consuming a large amount of medium time.

2) Reliability Loss: Multi-rate wireless devices are inherently designed to deal with changes in connectivity due to mobility and interference. The devices provide multiple link speeds to accommodate fluctuations in link quality. In 802.11b, as two nodes move in opposite directions, the auto rate protocol

TABLE V

Temporal Fairness Throughput Results

	Packet Fairness RBAR (Mbps)	Temporal Fairness OAR (Mbps)
11.0 Mbps Link	0.896	3.533
1.0 Mbps Link	0.713	0.450
Total	1.609	3.983

will gracefully reduce their link speeds from 11 Mbps down to 1 Mbps before they are finally disconnected.

Minimum hop path route selection has a tendency to choose routes that utilize the lowest link speed, leaving the auto rate protocol no flexibility in dealing with channel quality fluctuations. As a result, routes are often established between nodes that are on the fringe of connectivity. This occurs when nodes are able to receive broadcast transmissions, but data/ack packets are unable to be successfully delivered. Routing broadcasts can be extremely small in size while the data packets typically occupy the full frame size, making them more susceptible to the bit error rate (BER) of a low quality. This tendency is even further exaggerated by the way 802.11 handles broadcast transmissions as opposed to unicast transmissions. Unicast transmissions require a full RTS-CTS-DATA-ACK exchange for successful delivery, which is more likely to be disrupted by a low quality channel. The end result is that broadcasts can often be delivered even when symmetric communication is not possible.

B. Shortest Widest Path

Shortest widest path is a commonly used routing criteria in wired networks. This technique selects the shortest path from the set of paths that have the fastest bottleneck link. In wired networks, the total throughput of a path is directly related to the speed of the bottleneck link since each link in the path operates independently. Thus the shortest widest path is often used when high throughput is required.

There are a number of reasons why this path selection technique is not appropriate for wireless ad hoc networks. In wireless networks, individual links do not operate independently of one another. Individual transmissions affect a large area, and compete for medium time with both other transmissions along the same path and any transmission in the same geographical area. In addition, the shortest widest path does not consider the speed of links other than the bottleneck even though these links my affect the bottleneck link.

Consider the example shown in Figure 8. In this example, shortest widest path selection algorithm would select the two 1 Mbps links as the least cost path, despite the fact that the other path is almost 50% faster. It would consider each of the two paths equal with regard to the throughput they provide since they each have equal bottleneck links, therefore it would make

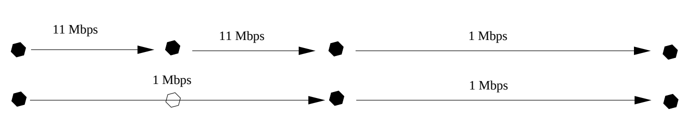

Fig. 8. Shortest Widest Path Selection

its selection based on which path contains the fewest hops. In this specific case, and in general when the bottleneck is a lower rate link, shortest widest path makes its selection based on hop count, and thus suffers from the same problems that apply to the minimum hop path.

Another example in which the shortest widest path makes an inappropriate trade-off is in a situation where a large number of high-rate links form one path, and a small number of lower rate links form another. Shortest widest path will select the long path because it has a higher bottleneck speed. However, the effective throughput drops as the path length increases. Therefore, it would be better to take the short path. For example, shortest widest path would select a path consisting of ten 11 Mbps links instead of a single 5.5 Mbps link. As illustrated in Figure 6, a single 5.5 Mbps link yields greater than five times the throughput of ten 11 Mbps links.

VI. General Model and Optimality Analysis

There is some ambiguity in the literature regarding what constitutes an optimal solution for the routing problem in multi-hop wireless networks. One of the main reasons for this is the difficulty inherent in modelling the complex environment of wireless multi-hop networks. We provide a model that captures many of the effects discussed above.

A. General Model of Attainable Throughput

In this work, we ignore packet scheduling issues and consider a steady-state flow model. In this model, each network edge may be fractionally shared by several flows; however, the sum of shares cannot exceed 100%. Our model of the wireless network is defined by a *transmission graph* and *interference graph*.

The *transmission graph* is defined as $G(V, E, \rho)$. V is defined as the set of nodes in the network. A transmission edge $(u, v) \in E$ if node u is capable of transmitting to node v. ρ is a function that assigns a transmission rate to each transmission edge $\rho : E \rightarrow R^+$. $\rho(e) = \hat{\rho}$ where $\hat{\rho}$ is the maximum flow rate obtainable over edge e when no other traffic exists in the network. $\hat{\rho}$ should take into account any sources of overhead such as contention, headers, and multiple frame exchanges, and represents the “real” capacity of edge e. In this general definition, the transmission graph may be directed, and the transmission rate in the reverse direction of a bi-directional edge may be different than that in the forward direction. This is possible in real wireless networks because of different node configurations and asymmetric channel effects.

The *interference graph* is defined as $G(\tilde{V}, \tilde{E})$. We define the vertices of the interference graph to be the edges of the transmission graph, so $\tilde{V} = E$. An edge in the interference graph represents the interaction between packets transmitted on nearby transmission edges. $((a, b), (c, d)) \in \tilde{E}$ if $(a, b), (c, d) \in E$ and if a transmission on (a, b) interferes with a transmission on (c, d).

In the general case, modelling the interference graph of an arbitrary network may be quite difficult due to complex propagation effects caused by obstacles and reflections. However, in the open space simulation configuration used in this, and many other papers, modelling the interference graph is much simpler. In this open space environment, the interference graph includes “edges” between each possible transmission edge, and all other transmission edges with an endpoint within carrier sense range of one of the transmission edge’s endpoints. This roughly corresponds to everything within a two hop neighborhood of a transmitting node.

Given the interference graph, we can define the interference neighborhood of any given edge (u, v) as follows.

$$\chi(u, v) = (u, v) \cup ((x, y) : ((x, y), (u, v)) \in \tilde{E})$$

Consider a set of i flows, where each flow ϕ_i originates from source s_i and is sinked by receiver r_i. Without loss of generality, we can represent each flow as a sum of path flows.

$$\phi_i = \sum_j \phi_{ij}$$

Each path flow ϕ_{ij} exists only on π_{ij}, where π_{ij} is a path from s_i to r_i in the transmission graph. In other words, $\phi_{ij}(x, y)$ equals the magnitude of the path flow $|\phi_{ij}|$ if the edge lies on its path, $(x, y) \in \pi_{ij}$, or zero otherwise.

With this setup, we can now specify a flow constraint that captures the phenomena discussed above. For each edge (u, v) in the transmission graph, the sum of the fractional shares used by all flows in the interference neighborhood of (u, v) must be less than or equal to 100%. This is a more complicated version of the classic edge capacity flow constraint.

$$\sum_{(x,y)\in\chi(u,v)} \sum_{i,j} \left(\frac{\phi_{ij}(x, y)}{\rho(x, y)} \right) \leq 1$$

In this general case, Linear Programming (LP) methods are required to achieve an optimal thruput solution. Opportunity-cost based approximations are possible in both the off-line case [21] (all connections are known ahead of time) and in the online case [22], [23]. Single path solutions are even harder to achieve as they require integer LP approaches.

B. Optimal Routing Assuming A Complete Interference Graph

Consider the special case of the general model where the interference graph is a clique (completely connected graph), i.e. each node can carrier sense each other node. In this special case, the constraint can be simplified since the interference neighborhood of any edge $\chi(u, v)$ is the same and consists of every edge in the transmission graph. Therefore we can rewrite the general flow constraint.

$$\sum_{(x,y)\in E} \sum_{i,j} \left(\frac{\phi_{ij}(x, y)}{\rho(x, y)} \right) \leq 1$$

We can reverse the order of summation.

$$\sum_{i,j} \sum_{(x,y)\in E} \left(\frac{\phi_{ij}(x, y)}{\rho(x, y)} \right) \leq 1$$

We can also decompose $\phi_{ij}(x, y)$ by moving its magnitude out of the inner sum, and changing the inner sum to include only non-zero terms.

$$\sum_{i,j}\left(|\phi_{ij}|\cdot\sum_{(x,y)\in\pi_{ij}}\left(\frac{1}{\rho(x,y)}\right)\right)\leq 1$$

Since $\rho(x,y)$ was defined as the real capacity of transmission edge (x,y), we can define the transmission time used by a unit of flow on this edge to be the inverse of this capacity.

$$\tau(x,y)=\frac{1}{\rho(x,y)}$$

Thus the final constraint equation becomes

$$\sum_{i,j}\left(|\phi_{ij}|\cdot\sum_{(x,y)\in\pi_{ij}}\left(\tau(x,y)\right)\right)\leq 1$$

In other words, the flow over each sub path consumes a certain fraction of the capacity. The sum of these fractions must be less than one. The fraction consumed by each sub path is equal to the amount of flow on that path times the sum of the transmission times along that path. The magnitude of flow on a sub path, $|\phi_{ij}|$, will be maximized when the sum of the transmission times along that path, $\sum_{(x,y)\in\pi_{ij}}\tau(x,y)$, is minimized. Therefore, a routing protocol that selects paths that minimize the sum of the transmission times maximizes the flow along those paths. Also, it is only necessary for each flow to have a single sub path that minimizes the sum of the transmission times, because any other sub paths will be at best equivalent to the minimum, and thus offer no additional flow capacity. Even if a flow does not use its maximum available capacity, minimizing the path transmission time minimizes the flow's consumption of the common network resource and allows other flows to increase. Therefore we can state the following.

Theorem 1: In the case of a complete interference graph in the stated multi-rate ad hoc wireless network model, a routing protocol that chooses a path that minimizes the sum of the transmission times minimizes network resource consumption, and maximizes total flow capacity.

VII. Medium Time Metric

We propose a *medium time metric* (MTM) that is designed to allow any shortest path routing protocol to find throughput optimal routes assuming full interference. The MTM assigns a weight to each link in the network that is proportional to the amount of medium time used by sending a packet on that link. The weight of any given path is thus a sum that is proportional to the total medium time consumed when a packet traverses the whole path. As a result, shortest path protocols that use the medium time metric find paths that minimizes the total transmission time.

Our approach to routing in multi-rate networks is derived from the above optimal solution under the full interference assumption. We discuss the validity and effects of this assumption later in this section. The complexity present in the general solution is caused primarily by the overlapping neighborhood constraints. This means that a general optimal algorithm must monitor the medium time utilization at every node in the network, and disseminate that information in order to aid routing decisions.

Under this assumption, the behavior of the optimal solution becomes much less complex. It is no longer possible to route around congested areas of the network, as any network congestion affects the entire network. Thus the paths selected by the optimal protocol no longer depend on the level or distribution of traffic. Also, it is now unnecessary for the optimal protocol to use multiple paths simultaneously. If there are multiple paths available that minimize the transmission time, the optimal protocol may choose any one of them at random and then use that path exclusively. Using additional paths offers no advantage as they would experience the exact same congestion. Thus a protocol that minimizes the total consumed medium time is throughput optimal given the full interference assumption.

A. Computing Link Weights

Our medium time metric states that paths that minimize the total consumed medium time should be selected. In order to accomplish this using existing shortest path protocols, we must assign a weight to each link that is directly proportional the medium time consumed by sending a packet across that link. The initial obvious solution is to use weights that are inversely proportional to the rate of the link. Using this scheme, if an 11 Mbps link was assigned a weight of 1, then a 1 Mbps link would be assigned a weight of 11 (see Table VI).

This inverse rate scheme turns out to be exactly the same as the default weight scheme suggested by Cisco for use with OSPF. However, in wired networks the MTM has none of the optimal characteristics that we argue it does in wireless networks. This is because the simplifying assumption that all transmissions interfere is completely invalid in wired networks; transmissions on one wire are completely isolated from transmissions on other wires. It is likely that Cisco recommends an inverse rate scheme because it represents a reasonable trade-off between high capacity and short paths.

However, we find that inverse rate weights do not accurately predict the amount of medium time consumed when sending a packet because they because they do not accurately represent an 802.11b packet transmission exchange. In 802.11b a packet is typically transmitted using a four frame MAC level exchange (RTS, CTS, DATA, ACK). Much of this exchange takes place at the 1 Mbps base rate, so a large nearly constant amount of medium time is consumed by per packet MAC overhead regardless of the actual link rate. This overhead becomes a large fraction of the total consumed medium time at the higher rates, because the actual data transmission time becomes small (see Figure 3). This overhead is why two nodes never achieve anywhere close to 11 Mbps of real throughput over an 11 Mbps link. For example, inverse weights would select a path of ten 11 Mbps links over a single 1 Mbps link. However, as shown in Section IV, a 1 Mbps link is clearly faster (and therefore consumes less medium time) than ten 11 Mbps links.

The almost fixed amount of medium time overhead caused by 802.11b introduces a dependency on packet size into our protocol. For example, the transmission time of a small packet will be dominated by the MAC overhead and will be almost the same regardless of the link rate. The implication of this phenomena is that the medium time metric would ideally use different link weights for each different packet size. This should be fairly easy to implement in link state protocols because they already have the topology information necessary to compute alternate routes using different sets of weights. However, this would be much more difficult for distance vector protocols, which require additional communication overhead for each additional set of weights. While it may be worth while in some networks to track more than one set of weights, usually the bulk of data transferred in a given network is of a single size.

An implementation of the MTM for a distance vector protocol should be tuned for the dominant packet size used by the network. This is accomplished by using link weights that are proportional to the medium time used by packets of the tuned size. These tuned weights represent the best trade-off point between short low-rate paths and long high-rate paths for packets of the tuned size. Packets that are much larger than the tuned size may have been better off traversing a longer path with even higher rate links. Similarly, packets much smaller than the tuned size may be better off taking paths that are shorter but with lower rate links. Performance of the MTM should not be significantly affected by transmissions of packets larger and smaller than the tuned size as long as those packets do not consume a large fraction of the total medium time.

In this work, the tuned packet size was chosen to correspond to a 1500 byte IP packet. This size is representative of the majority of the data transferred by the Internet [24] and corresponds to the standard Ethernet maximum transferable unit (MTU) [19]. This size was chosen over the larger native MTU of 802.11b (2314 bytes) because wireless networks today are mostly used to provide mobile access to LAN and Internet resources. In this environment, packets that flow over fixed links as well as wireless links would be limited to a 1500 byte path MTU. Purely peer to peer wireless networks would be free to use the native MTU, and could gain an additional measure of throughput due to the increased ratio of data to overhead in each packet.

Table VI shows the expected medium times, and corresponding proportional weights, for each rate computed according to the 802.11b standard specifications. These weights are significantly different then the inverse weights. The times are calculated assuming a full RTS, CTS, DATA, ACK exchange. All information is sent at the base 1 Mbps rate except for the contents of the data and acknowledgement frames, which are sent at the chosen link rate. These computed times also include an estimate of the time spent backing off during contention. We used the value of half the minimum contention window size multiplied by the slot time (310 μsec). This estimate was derived from the average time spent in the single sender case, but should

TABLE VI

RATE BASED LINK WEIGHTS

Link Rate (Mbps)	Inverse Weights	1500 byte packet μsec	1500 byte packet MTM Weights
11.0	1.00	2542	1.00
5.5	2.00	3673	1.44
2.0	5.50	7634	3.00
1.0	11.00	13858	5.45

function sufficiently for multiple senders. When we have an increased number of senders contending, the average idle medium time should decrease dramatically because the time spent for any particular packet is the minimum of all the senders random back offs. However, the probability of a collision also increases so the average time wasted while performing contention should not change as much as we might expect.

Even though a large number of acknowledgement packets are present in the network when TCP is used, the time total consumed by these packets is small in comparison to the data. This is particularly true when the delayed acknowledgement option of TCP is used which effectively halves the number of acknowledgements. OAR further reduces the proportion of time consumed by acknowledgement packets sent at high-rate by amortizing much of the contention and control overhead over several packets.

Since the OAR protocol significantly changes the MAC layer packet exchange, the expected medium time consumed by a packet at a given rate changes significantly. Thus in networks where OAR, or a significantly different MAC exchange, is used, different MTM weights must be calculated to match the change in consumed medium time.

B. Advantages

The medium time metric has several advantages over other possible routing strategies. One of its primary advantages is its simplicity. As a shortest path metric, it can be incorporated into existing distance vector or link-state protocols. The majority of existing wireless ad hoc routing protocols fall into these categories (AODV, DSR, OLSR, DSDV). It would be much more difficult to incorporate the MTM into protocols that use routing strategies other than shortest path, such as TORA [25].

The medium time metric also sidesteps the most serious problems exhibited by the optimal solution under the general model. MTM protocols only need to track changes in link rates as opposed to changes in utilization. This results in drastically lower protocol overhead. Also, there is no danger of route oscillation because MTM routes do not depend on traffic patterns, There is no danger of disrupting higher level protocols such as TCP due to out of order packet delivery because the MTM selects a single path.

While our medium time metric avoids the problems of the optimal solution under the general model, many of that solution's desirable characteristics are preserved. The minimum transmission time paths selected according to the MTM minimize the total interference created in the network. A minimum transmission time path represents the most efficient path, with respect to medium time, over which to deliver a packet. Choosing the most efficient paths also has the the effect of increasing the total network capacity, thus yielding high total throughput.

When compared with min-hop paths, MTM paths increase efficiency by choosing more medium or high-rate links in order to minimize the number of low-rate transmissions that consume a large amount of medium time. When compared with shortest widest paths, MTM paths increase efficiency by avoiding taking long paths of high-rate links in favor of shorter paths of medium or even low-rate links.

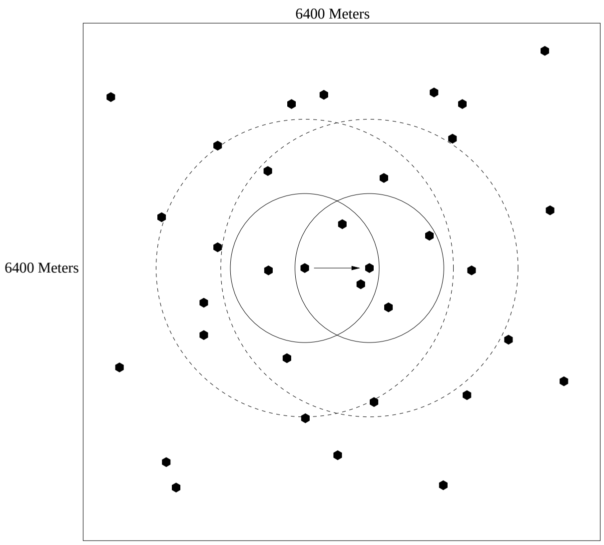

Fig. 9. Interference Range in a 6400 m x 6400 m Network

Another interesting property of MTM paths is that since they naturally avoid low-rate links, they exhibit some of the properties of signal stability based routing protocols. Nodes connected by a high-rate link must a considerable distance before the link breaks. As the nodes move further apart, the auto rate protocol reduces the link speed. As a result, proactive routing protocols, which continually update their paths based on the MTM will naturally avoid path failures by continuously switching to higher rate links.

C. Discussion

The derivation of the medium time metric involves a simplifying assumption. The obvious question is how accurate is this assumption, and what impact will it have on the performance of our solution. In real networks the interference graph is primarily determined by the carrier sense range (see Figure 5). While the carrier sense range is not infinite, it is significantly greater than even the maximum transmission range. In a small network (up to 1783 meters in diameter), all nodes are indeed in carrier sense range of each other, and thus the assumption holds perfectly. Even in a much larger area, such as the one shown in Figure 9, a single transmission can still interfere with a large proportion the network. The performance of the MTM in comparison to the optimal should not be significantly degraded until the carrier sense range becomes small in comparison to the path length.

For example, as the path length grows longer, then the optimal can achieve additional throughput by sending packets over several “disjoint” paths. These paths must be at least a carrier sense range away from each other to avoid interference, but since they must eventually converge at the the end points of the path, they can never be completely disjoint. This unavoidable interference around the source and destination means that two nodes must be further away than twice the carrier sense range before any throughput gains can be achieved by using multiple paths. Also, the additional throughput will be small, due to the long path lengths necessary to avoid interference from other adjacent paths.

In a study of the capacity of ad hoc wireless networks [26], the authors show that these networks only scale if the traffic patterns remain local. If we assume local traffic patterns, such as when every node accesses the nearest available Internet connected node, the MTM assumption may still be valid even in large networks. This is because even though the total network diameter is large, communication actually occurs in much smaller sub-network cells formed around the Internet connected nodes.

If the traffic patterns are not local in a large network, even an optimal routing algorithm will achieve low throughput. This is because physically long paths require many hops, which have low path throughput and consume resources all across the network.

It is important to note that link rates by definition change faster than link connectivity. As a result, some routing protocols may consume more overhead when using MTM when compared with min hop. However, the degree of extra overhead, if any, is completely dependent on the routing protocol and its implementation. We also point out that since distance is the dominant factor that determines the link rate, even in the worst case, the MTM metric should only change a constant amount more than connectivity. This is significantly better than any routing solution that is traffic sensitive, as traffic loads change much faster than either link rates or link connectivity.

Typically, the MTM selects paths that have a greater number of hops than the minimum. While these higher rate hops consume less total medium time than the minimum number of hops, the increased number of senders could cause other detrimental effects.

The increased number of senders creates higher contention for the medium. If the efficiency of the medium access control protocol degrades significantly as demand increases, then the efficiency of the medium time metric will also degrade. The authors of [27] present a detailed analysis of TCP performance in multi-hop wireless networks running the 802.11 MAC. The paper shows that an entire multi-hop path collectively operates similarly to a single RED queuing gateway. The probability of a packet being dropped due to contention along the path increases linearly with increased offered load beyond a threshold. They also note that the maximum drop probability is determined by the total amount of contention experienced along the path. The authors show that nearly all dropped TCP packets are a result of contention losses as opposed to buffer overflows. Also, they note that unmodified TCP generally grows its congestion window too high for wireless networks, causing reduced throughput. MTM causes increased contention by increasing the number of senders, which should cause the probability of a packet drop to increase. The resulting higher drop probability is actually a desirable effect as it should help reduce the oversized TCP congestion window, in fact the increased drop probability should provide a similar effect as the LRED solution proposed by the authors.

Any potential performance decreases due to increased contention would also be reduced when using the OAR protocol. This is because the multi-packet burst of OAR effectively reduces the total time spent in contention when high-rate links are used.

An additional result of increased hop count is that there are more interface queue buffers along the path a packet must traverse. This increased amount of buffering could lead to an increase in end-to-end latency when the network is congested. While trading end-to-end latency for increased throughput is completely appropriate for bulk data transfer applications, which is not the case for delay sensitive traffic. Priority queues should be used on the intermediate nodes regardless of the routing metric used. This eliminates the need to wait in line at multiple buffers. It is also important to realize that although a min hop path may seem appropriate for delay sensitive traffic, it may actually take longer to deliver a packet over the min hop path than an MTM path. This is because it takes much longer for a non-zero sized packet to be delivered across across a low-rate link as opposed to a high-rate link.

Many types of delay sensitive packets, such as Telnet traffic, use relatively small packet sizes. Small delay sensitive packets would benefit from MTM routes tuned for small packet sizes. An implementation of MTM which tracks multiple packet sizes would also be effective.

Routing protocols that use the medium time metric choose paths that minimize the total consumed medium time. We have argued that these paths should yield significant throughput gains when compared with minimum hop paths. However, this assumes that a path exists that utilizes less medium time than the minimum hop path. This may not be the case. Whether a better MTM path exists depends solely on the current network topology. In general, the likelihood of there existing a smaller medium time path increases as the density of the network increases.

When the density of the network is low, the topology becomes sparsely connected. This yields few choices for routing protocols to select from. In this situation, MTM and min hop will tend to pick the same path. Conversely, as the network density increases, the abundance of nodes creates a dense, heavily interconnected topology. Routing protocols are provided with a multitude of paths from which to choose. This large number of choices allows the natural tendencies of each metric to be expressed fully.

We have constructed a simple experiment designed to illustrate the relationship between density and the performance of the MTM. A variable number of nodes are randomly placed along a straight line path of fixed length. A single UDP flow is setup between the source and destination, which are placed at opposite ends of the line. Figure 10 shows the relative throughput of the MTM and min hop routing protocols as the number of nodes and the line length are varied. The vertical axis shows the percent increase in achieved throughput over the min hop path when using the MTM. The horizontal axis shows the normalized density of the topology. We define the *normalized density* as the average number of nodes within the maximum transmission range of a given node. In this line case, density is given by the following equation.

$$density = \frac{2 \cdot 786 meters \cdot nodes}{length}$$

The results show a clear relationship between node density and increased throughput. As expected, at low densities we

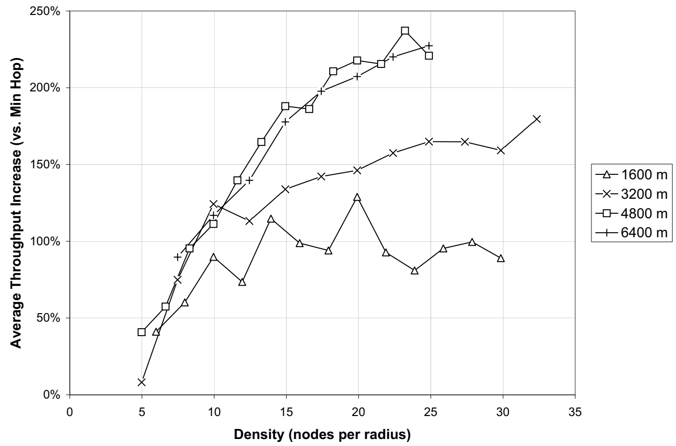

Fig. 10. Average throughput increase of MTM over min hop along a randomized straight line path.

see low increases as both the MTM and min hop metric pick nearly the same path. As the density increases, we see the full potential of the MTM revealed. The MTM path yields greater than three times (+200%) the throughput of the min hop path with the higher densities and longer path lengths. Longer paths yield more increased throughput than shorter paths because the MTM path utilizes the extra medium time available in long paths (from spatial reuse) much more efficiently than the min hop path.

VIII. Implementing the Medium Time Metric

A. On-Demand Protocols

Although the results presented in this work are based on a modified version of DSDV which is a pro-active routing protocol, the methodology is applicable to on-demand protocols as well. On-demand routing protocols, such as AODV [5] and DSR [6] initiate a route discovery process only when packets need to be routed, in an attempt to minimize the routing overhead. By utilizing the weight scheme that we present, on-demand protocols would discover the least cost path between the source and the destination.

By amortizing the cost of route discovery over the period of time in which the route is valid, one can argue that if two routing algorithms pay the same cost to discover a route, then the algorithm which chooses routes that last for a longer duration of time pays less overhead. We will refer to a routes ability to withstand changes due to mobility as the *elasticity* of the route.

An important distinction between on-demand protocols and pro-active protocols is that on-demand protocols typically maintain a discovered route until it fails. This presents a number of issues with regard to the applicability of the technique we present. If on-demand protocols used our weight scheme to discover paths, the resulting paths would be more elastic and provide connectivity for a longer duration of time under mobility. The nodes selected would be less likely to be on the fringe of the communication range. The problem with using this technique with on-demand protocols is that although the path may provide connectivity for a long duration of time, mobility will cause the performance of the path to deteriorate. Solutions to this problem typically involve a trade-off between the path performance and the routing overhead.

B. Modifying DSDV

In order to allow the DSDV routing protocol [4] to use the medium time metric, two major modifications were needed.

1) Rate Metric: The first major modification is to use a distance metric based on rate dependent weights as opposed to hop count. Instead of adding one to the metric advertised by a neighbor, a node adds a weight indexed by the current communication rate with that neighbor.

The current communication rate is determined based on information passed up from the MAC layer. Our implementation assumes that an estimate of the current rate is passed up the stack along with each received update packet. An integrated event based scheme where the MAC notifies DSDV whenever it updates its estimate (e.g. a neighbor was transmitting data or beacons) could improve performance by allowing DSDV to react faster to rate changes. However, this tight integration is more difficult to implement than passing a small amount of extra information up the stack along with a received packet.

In order to employ our medium time metric, we use weights that are directly proportional to the medium time consumed by transmitting packets on a link of the corresponding rate. Since the metric field of a DSDV update packet is a single byte, we must choose small integer weights in order to avoid overflowing it when long paths are present. Making the weights small in order to accommodate long paths reduces the precision of the weights and could compromise the path selection accuracy, especially for long paths. As a compromise, we normalize the calculated medium times so that the slowest rate corresponded to a weight of 25. This guarantees that all paths up to at least 10 hops can be represented. Paths of much greater than 10 hops can be represented as long as they consist of higher rate links.

2) Route Selection Oscillation: The second major modification dampens route selection oscillation. This oscillation is a result of DSDV's route selection criteria and update propagation behavior. This route oscillation can occur even in a static network because updates are propagated asynchronously. The problem particularly affects multi-rate networks because updates are likely to be first heard across long range low-rate links.

Our solution to the problem is similar to the existing DSDV multiple update suppression scheme. We modify DSDV's route selection criteria to wait an additional settling time before installing a route with a newer sequence number but a greater metric than the current route. This installation delay greatly dampens the route selection oscillation.

IX. Simulation Results

The purpose of this section is to evaluate the techniques proposed in this paper in a full simulated network environment. We explored the throughput gains provided by both our proposed medium time metric (MTM) and the temporally fair opportunistic auto rate (OAR) protocol over the traditional minimum hop (min hop) metric and the packet fair receiver based auto rate protocol (RBAR).

As opposed to the small simple setups presented in previous sections of this paper, the simulations here are designed to be representative of real networks. Several parameters of the simulation setup are fixed for all the simulations in this section, and some are varied in order to illustrate their effects on the throughput gains.

Fixed Parameters. The wireless physical parameters given in Section III-B are used. In every simulation, a random way-point mobility model is used. Our simulations are setup for high mobility: the maximum speed is set to 20 meters per second and the pause time is set to zero seconds. In order to emulate a network under high load, we setup 20 flows of TCP traffic. We use the delayed acknowledgement option of TCP in order to reduce the medium time consumed by TCP acknowledgements. Each average gain result is computed from the gains in at least 25 random scenarios. Each scenario is created using a random number seed that generates the initial node placement and mobility pattern. The gains are computed by simulating each of the four protocol combinations (RBAR & min hop, RBAR & MTM, OAR & min hop, and OAR & MTM) in the exact same scenario, and then dividing the resulting total throughput by the base combination (RBAR & min hop). The base combination is representative of both the standard 802.11 MAC fairness model and the metric used by the majority of existing ad hoc routing protocols. This technique of computing gains prevents scenarios with high throughput from skewing the final average. Min hop results are obtained by using the standard DSDV protocol. MTM results are obtained using the modifications to DSDV specified in Section VIII-B. MTM link weights are tuned to match both the TCP traffic, which carries a 1460 byte payload in these simulations, and the selected auto rate protocol (RBAR or OAR).

Varying Parameters. The primary variable examined in this section is node density. The effect of node density on throughput gains was shown at the end of Section VII-C, but only in a simpler one dimensional line case. The central question this section hopes to answer, is how many nodes are required to reach the point where the MTM metric can increase throughput by selecting better paths. It is clear that almost any reasonable routing metric will achieve similar performance when the network density is low because the number of available paths to choose from is limited. Since we have defined the *normalized density* as the average number of nodes within the maximum transmission range of a given node, the density in these scenarios is a function of both the number of nodes in the simulation and the total area of the simulation topology. We present simulation results for 60, 100, 150 and 193 nodes in 2400 meter by 2400 meter and 3200 meter by 3200 meter sized topologies.

Results. Figure 11 shows the average throughput gains with respect to the throughput of the RBAR & min hop combination. The average gains of the RBAR & MTM combination represent the throughput increase achieved by our proposed medium time metric under the standard packet fairness model of 802.11. As expected, we see a clear increasing trend in average gain as the density increases. Even at the lowest node density, MTM provides a modest 18% average increase. At the highest simulated density, we see a more substantial 56% average increase. The gains should continue to climb with even higher densities until

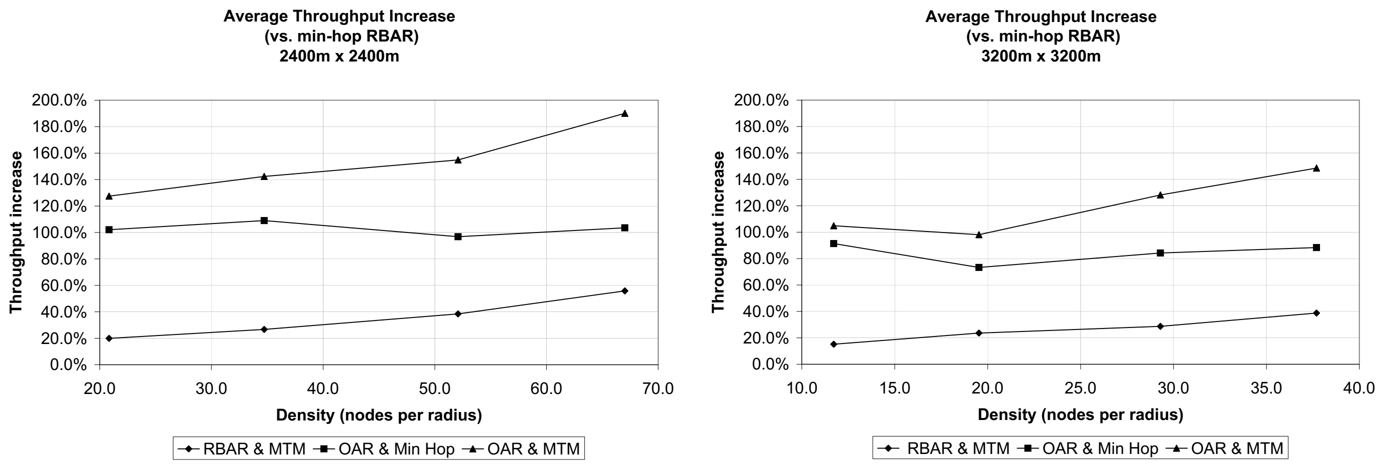

Fig. 11. Random Motion Average Throughput Gains

a plateau is reached. Due to the increased degree of freedom in comparison to the line case, the plateau should not occur until high densities are reached.

For reference, the two sizes of simulations used in [28], a performance comparison of AODV and DSR, were 1500 meters by 300 meters with 50 nodes and 2200 meters by 600 meters with 100 nodes. Given the 250 meter nominal range used in this comparison, these simulations have the normalized densities of 21.8 and 14.9 respectively.

The average gains of the OAR & min hop combination show the throughput increase produced by the OAR protocol without changing the routing metric. As shown in the results, OAR provides quite a substantial boost in total network throughput. The gains provided by OAR come from two sources: increased overall network efficiency due to the increased proportion of time spent sending at high rates, and reduced MAC overhead due to amortization over a multiple packet burst. As a result, the OAR gains are relatively constant with respect to the node density. This experiment illustrates that the OAR protocol should be used in high throughput multi-rate networks even if min hop is used as the routing metric.

Our analysis of the wide variety of phenomena that affect the throughput in multi-rate ad hoc wireless networks suggests that our proposed medium time metric and the OAR protocol should function well together. The MTM generally selects paths with higher rate links than the min hop, and thus gains an increased benefit from the reduced MAC overhead of high-rate links provided by OAR. Since MTM picks a greater number of high-rate links, it receives less benefit from the temporal fairness property of OAR, but is still helped in the case where paths with fast links are not available. The simulation results show that OAR and MTM do indeed function well together. The contribution of the MTM introduces the same kind of dependance on density that we saw in the pure MTM results. In the most dense simulated case, the total network throughput is almost tripled on average. These massive throughput gains lend support to the validity of both the analysis and solution presented in this paper.

X. Conclusion

In this work we have shown that minimum hop protocols tend to select paths with long slow links. As a result, these paths have low effective throughput and increase total network congestion. In addition, these paths are likely to contain long links that result in low reliability.

We have presented an improved technique for route selection in multi-rate ad hoc wireless networks. The medium time metric is proportional to the time it takes to transmit a packet on a given link. This metric selects paths that have the highest effective capacity. We have also shown the optimality of this technique by presenting a formal theoretical model of the attainable throughput of multi-rate ad hoc wireless networks.

Our simulation results show an average throughput gain of 20% to 60%, depending on network density, over traditional minimum hop route selection in 802.11b networks. By combining the MTM with the Opportunistic Auto Rate (OAR) protocol, an increase of 100% to 200% is obtained over the traditional route and rate selection techniques. Our results demonstrate the importance of inter-layer communication in ad hoc routing protocol design.

REFERENCES

[1] *IEEE Std 802.11b-1999*, http://standards.ieee.org/.

[2] *IEEE Std 802.11a-1999*, http://standards.ieee.org/.

[3] *Draft ESTI EN 301 893 version 1.1.1: Broadband Radio Access Networks; HIPERLAN Type 2*, http://www.etsi.org/.

[4] Charles E. Perkins and Pravin Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications*, 1994.

[5] Charles E. Perkins and Elizabeth M. Royer, *Ad hoc Networking*, chapter Ad hoc On-Demand Distance Vector Routing, Addison-Wesley, 2000.

[6] David B. Johnson, David A. Maltz, and Josh Broch, *DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks. in Ad Hoc Networking*, chapter 5, pp. 139–172, Addison-Wesley, 2001.

[7] A. Laouiti P. Muhlethaler a. Qayyum et L. Viennot T. Clausen, P. Jacquet, "Optimized link state routing protocol," in *IEEE INMIC*, Pakistan, 2001.

[8] Douglas S. J. De Couto, Daniel Aguayo, Benjamin A. Chambers, and Robert Morris, "Performance of multihop wireless networks: Shortest path is not enough," in *Proceedings of the First Workshop on Hot Topics in Networks (HotNets-I)*, Princeton, New Jersey, October 2002, ACM SIGCOMM.

[9] R. Dube, C. Rais, K. Wang, and S. Tripathi, "Signal stability based adaptive routing (ssa) for ad hoc mobile networks," February 1997.

[10] Henrik Lundgren, Erik Nordstrom, and Christian Tschudin, "Coping with communication grey zones in ieee 802.11b based ad hoc networks," in *WoWMoM 2002*, September 2002.

[11] *ANSI/IEEE Std 802.11, 1999 Edition*, http://standards.ieee.org/.

[12] A. Kamerman and L. Monteban, "WaveLAN-II: A high-performance wireless lan for the unlicensed band," in *Bell Labs Technical Journal*, Summer 1997, pp. 118–133.

[13] Anand R. Prasad and Henri Moelard, "WaveLAN-II system design note 225: Enhanced data rate control," March 1999.

[14] Gavin Holland, Nitin H. Vaidya, and Paramvir Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Mobile Computing and Networking*, 2001, pp. 236–251.

[15] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," September 2002.

[16] "The network simulator - ns2," http://www.isi.edu/nsnam/ns/.

[17] "Rice networks group," http://www-ece.rice.edu/networks/.

[18] "Orinoco wireless networks," http://www.orinocowireless.com/.

[19] *IEEE Std 802.3-2002*, http://standards.ieee.org/.

[20] F A Tobagi and L Kleinrock, "Packet switching in radio channels: Part ii - the hidden terminal problem in carrier sense multiple-access and the busy tone solution," in *IEEE Transactions on Communications*, 1975, pp. 1417–1433.

[21] T. Leighton, F. Makedon, S. Plotkin, C. Stein, E. Tardos, and S. Tragoudas, "Fast approximation algorithms for multicommodity flow problem," in *Proc. 23rd ACM Symp. on Theory of Computing*, May 1991, pp. 101–111.

[22] J. Aspnes, Y. Azar, A. Fiat, S. Plotkin, and O. Waarts, "On-line machine scheduling with applications to load balancing and virtual circuit routing," in *Proc. 25th ACM Symp. on Theory of Computing*, May 1993, pp. 623–631.

[23] Baruch Awerbuch, Yossi Azar, and Serge Plotkin, "Throughput competitive on-line routing.," in *Proc. 34th IEEE Symp. on Found. of Comp. Science*. Nov. 1993, pp. 32–40, IEEE.

[24] Cooperative Association for Internet Data Analysis (CAIDA), "Analysis of NASA Ames Internet Exchange Packet Length Distributions," http://www.caida.org/.

[25] Vincent D. Park and M. Scott Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *INFOCOM (3)*, 1997, pp. 1405–1413.

[26] Jinyang Li, Charles Blake, Douglas S. J. De Couto, Hu Imm Lee, and Robert Morris, "Capacity of ad hoc wireless networks," in *Proceedings of the 7th ACM International Conference on Mobile Computing and Networking*, Rome, Italy, July 2001, pp. 61–69.

[27] Zhenghua Fu, Petros Zerfos, kaixin Xu, Haiyun Luo, Songwu Lu, Lixia Zhang, and Mario Gerla, "On tcp performance in multihop wireless networks," .

[28] C. Perkins, E. Royer, S. Das, and M Marina, "Performance comparison of two on-demand routing protocols for ad hoc networks," in *IEEE INFOCOM*, 2000.