Simulating mobile ad hoc networks: a quantitative evaluation of common MANET simulation models

Calvin Newport

Dartmouth College Computer Science Technical Report TR2004-504

June 16, 2004

Abstract

Because it is difficult and costly to conduct real-world mobile ad hoc network experiments, researchers commonly rely on computer simulation to evaluate their routing protocols. However, simulation is far from perfect. A growing number of studies indicate that simulated results can be dramatically affected by several sensitive simulation parameters. It is also commonly noted that most simulation models make simplifying assumptions about radio behavior. This situation casts doubt on the reliability and applicability of many ad hoc network simulation results.

In this study, we begin with a large outdoor routing experiment testing the performance of four popular ad hoc algorithms (AODV, APRL, ODMRP, and STARA). We present a detailed comparative analysis of these four implementations. Then, using the outdoor results as a baseline of reality, we disprove a set of common assumptions used in simulation design, and quantify the impact of these assumptions on simulated results. We also more specifically validate a group of popular radio models with our real-world data, and explore the sensitivity of various simulation parameters in predicting accurate results. We close with a series of specific recommendations for simulation and ad hoc routing protocol designers.

1 Introduction

It is difficult to perform an accurate evaluation of a mobile ad hoc network (MANET). In a perfect world, ad hoc network protocols would always be validated with extensive real world experimentation. The best way to predict the behavior of a network is to deploy it in a real environment, and then observe what happens. For obvious reasons, however, such experimentation is rarely done. As Zhang et al. point out, “running MANET systems in a non-trivial size is costly due to high complexity, required hardware resources, and inability [to test] a wide range of mobility scenarios” [ZL02]. This difficulty is supported by our own anecdotal experience. Conducting the outdoor experiment reported in this study required over two years of preparation by a team of more than ten researchers and student interns. In fact, Zhang et al. report that their literature search uncovered only a “few” real world systems that have *ever* been implemented, and none that have been tried on a scale beyond a dozen nodes [ZL02]. Our review of the MANET literature, two years later, confirms this observation.

Because of this difficulty of running real world experiments, it is clear that, at least for now, computer simulation will remain the standard for ad hoc network evaluation. However, this reliance on simulation demands in return a careful scrutiny of common simulation approaches. For simulation to be used as a meaningful evaluation technique, there must be a concerted effort to understand the models being used—including their specific characteristics, and their relative validity. This conclusion is supported by an increasing body of research that demonstrates that the outcome of wireless network simulation is quite sensitive to the underlying models. For example, in an experiment conducted by Takai et al. [TMB01], it is shown that altering parameters in commonly used radio models has a non-uniform effect on ad hoc protocol behavior, sometimes even reversing the relative ranking among protocols tested in the same scenario. The simulation results reported in this study similarly demonstrate dramatic changes in outcomes when different radio models are used (see Section 4.5). And a recent article in *IEEE Communications* warns that “An opinion is spreading that one cannot rely on the majority of the published results on performance evaluation studies of telecommunication networks based on stochastic simulation, since they lack credibility” [PJL02]. It then proceeds to survey 2200 published network simulation results to point out systemic flaws.

We of course do not suggest that there is one *right* answer to the question of simulation validity. Accordingly, we do not attempt to identify any one *right* model that always performs best. But we do, however, argue that simulation designers should explicitly address the assumptions made in their models, and the influence that these assumptions may have on simulation results. We believe that this approach will allow the MANET community to confidently find relevance for simulation outcomes beyond the specific simulator configuration in which a particular experiment was run.

With this in mind, we identify several specific questions that simulation designers should consider and strive to answer to ensure that their results are as meaningful as possible:

- What assumptions are made by the radio propagation model?
- How realistic are these assumptions?
- What is the effect of these assumptions on the results?
- Has the radio model been validated with experimental data, and if so, how does it perform relative to reality? For example, does the model tend to predict a higher rate of network connectivity than what is observed under experimental conditions? Does it exaggerate the maximum range of the network's radios? What is the significance of these variations for understanding the simulation results?
- What simulation parameters are used? How sensitive are the results to changes in these parameters? And how do the values used constrain the applicability of the results?

In this study, we detail the necessity of these questions, and use real experimental data combined with a wireless network simulator to explore answers as they apply to a group of commonly used radio propagation models. More specifically, we a) describe in detail the implementation and results of a large outdoor MANET routing experiment; b) identify the extent to which most ad hoc network researchers make common simplifying assumptions about the radio model used in simulation; c) use experimental data to quantitatively demonstrate that these assumptions are far from realistic; d) explore how these assumptions may lead to misleading results in ad hoc network simulation; e) validate the predictive power of our selected models against an experimental baseline; f) explore the role of important parameters in simulation results; and g) list recommendations for the designers of protocols, models, and simulators.

The results described in this study span over two years of work, including one published paper [LYN$^+$04], one technical report [KNE03] (with a revision currently under conference submission), and another conference paper in preparation that describes the large-scale outdoor MANET routing experiment, and analyzes the performance data. This thesis is the first complete synthesis of these various experiments into one comprehensive examination of accuracy in ad hoc network simulation. We hope that our use of real experimental results to ground our simulation analysis will make this work particularly useful for simulation designers, and provide an important contribution to the growing field of ad hoc networking research.

2 Outdoor routing experiment

As mentioned above, few MANET researchers conduct real-world experiments. The cost and complexity are prohibitive for most projects. In this section, however, we throw caution to the wind, and join a team that is up to the challenge of testing four popular ad hoc routing protocols in a dynamic outdoor environment. Specifically, research engineer Robert Gray organized a real-world routing experiment as part of a larger multi-disciplinary university research initiative led by Dartmouth Professor George Cybenko.[1] Gray worked on the scenario design with Susan McGrath, Eileen Entin and Lisa Shay, and the algorithms were implemented by Aaron Fiske, Chris Masone, Nikita Dubrovsky, and Michael DeRosa. Our role in this project is to gather and organize the data produced by the experiment and then provide a detailed comparative analysis of the results.

This description of real network behavior, in addition to being a stand-alone contribution to the research community, also forms an empirical baseline that aids our subsequent validation of ad hoc network simulation.

2.1 The algorithms

The outdoor experiment tests four algorithms. APRL, which stands for Any-Path Routing without Loops, is a proactive distance-vector routing protocol [KK98]. Rather than using sequence numbers, APRL uses ping messages before establishing new routes to guarantee loop-free operation. AODV, or Ad-hoc On-Demand Vector, is an on-demand routing algorithm—routes are created as needed at connection establishment and maintained thereafter to deal with link breakage [PR99]. ODMRP stands for the On-Demand Multicast Routing Protocol [LGC02]. For each multicast group, ODMRP maintains a mesh, instead of a tree, for alternate and redundant routes. ODMRP does not depend on another unicast routing protocol and, in fact, can be used for uni-

[1] http://actcomm.dartmouth.edu

cast routing. STARA, the System and Traffic Dependent Adaptive Routing Algorithm, is based on shortest-path routing [GK97]. It uses mean transmission delay instead of hop count as the distance measure.

2.2 The experiment

The outdoor routing experiment took place on a rectangular athletic field measuring approximately 225 (north-south) by 365 (east-west) meters. This field can be roughly divided into four flat, equal-sized sections, three of which are at the same altitude, and one of which is approximately four to six meters lower. There was a short, steep slope between the upper and lower sections.

Each Linux laptop[2] had a wireless card[3] operating in peer-to-peer mode at 2 Mb/s. This fixed rate made it much easier to conduct the experiment, since it obviated the need to track (and later model) automatic changes to each card's transmission rate.

To reduce interference from the campus wireless network, the experiment was conducted on a field physically distant from campus, and the cards were configured to use wireless channel 9 for maximum separation from the standard channels (1, 6 and 11). In addition, each laptop collected signal-strength statistics for each received packet.[4] Finally, each laptop had a Garmin eTrex GPS unit attached via the serial port. These GPS units did not have differential GPS capabilities, but were accurate to within thirty feet during the experiment.

Each laptop recorded its current position (latitude, longitude and altitude) once per second, synchronizing its clock with the GPS clock to provide sub-second, albeit not millisecond, time synchronization. Every three seconds, the *beacon service program* on each laptop *broadcast* a beacon packet containing the current laptop position (as well as the last known positions of the other laptops). Each laptop that received such a beacon updated its internal position table, and sent a *unicast acknowledgment* to the beacon sender via UDP. Each laptop recorded all incoming and outgoing beacons and acknowledgments in another log file. The beacons provide a continuous picture of network connectivity, and, fortunately, also represent network traffic that would be exchanged in many real MANET applications. Finally, every second each laptop queried the wireless driver to obtain the signal strength of the most recent packet *received* from every other laptop, and recorded this signal strength information in a third log.[5] Querying every second for all signal strengths was much more efficient than querying for individual signal strengths after each received packet.

These three logs provide all the data that we need to compare the performance of the four routing algorithms. The laptops automatically ran each routing algorithm for 15 minutes, generating random UDP data traffic for thirteen out of the fifteen minutes, and pausing for two minutes between each algorithm to handle cleanup and setup chores. The traffic-generation parameters were set to produce the traffic volumes observed in previously explored prototype situational-awareness applications [Gra00], approximately 423 outgoing bytes (including UDP, IP and Ethernet headers) per laptop per second, a relatively modest traffic volume. The routing algorithms produce additional traffic to discover or maintain routing information. Note that each transmitted data packet was destined for only a single recipient, reducing ODMRP to the unicast case.

Finally, the laptops moved continuously. At the start of the experiment, the participants were divided into equal-sized groups of ten each, each participant given a laptop, and each group instructed to randomly disburse in one of the four sections of the field (three upper and one lower). The participants then walked continuously, always picking a section different than the one in which they were currently located, picking a random position within that section, walking to that position in a straight line, and then repeating. This approach was chosen since it was simple, but still provided continuous movement to which the routing algorithms could react, as well as similar spatial distributions across each algorithm.

During the experiment, seven laptops generated no network traffic due to hardware and configuration issues, and an eighth laptop generated the position beacons only for the first half of the experiment. We use the data from the remaining thirty-two laptops, although when we simulate later, we use thirty-three laptops since only seven laptops generated no network traffic at all. In addition, STARA generated an overwhelming amount of control traffic, and though we exclude the STARA portion of the experiment from later analysis of radio behavior, we still present its data in the outdoor results sections that follow. The reason we exclude it later is because its unusual behavior makes it a poor reference point for the specific task of validating simulation results.

[2] A Gateway Solo 9300 running Linux kernel version 2.2.19 with PCMCIA Card Manager version 3.2.4.

[3] We used a Lucent (Orinoco) Wavelan Turbo Gold 802.11b. Although these cards can transmit at different bit rates and can auto-adjust this bit rate depending on the observed signal-to-noise ratio, we used an ad hoc mode in which the transmission rate was fixed at 2 Mb/s. Specifically we used firmware version 4.32 and the proprietary ad hoc "demo" mode originally developed by Lucent. Although the demo mode has been deprecated in favor of the IEEE 802.11b defined IBSS, we used the Lucent proprietary mode to ensure consistency with a series of ad hoc routing experiments of which this outdoor experiment was the culminating event.

[4] We used the `wvlan_cs`, rather than the `orinoco_cs`, driver.

[5] For readers familiar with Linux wireless services, note that we increased the IWSPY limit from 8 to 64 nodes, so that we could capture signal-strength information for the full set of laptops.

2.3 The results

To evaluate the relative performance of these algorithms we use the following four metrics: *message delivery ratio, communication efficiency, hop count, and end-to-end latency.* Combined, these measures provide a good understanding of the various factors involved in the different observed behaviors. In the sections that follow, we give a detailed definition of each metric, and compare the results for each algorithm.

Before proceeding, there are several important terms used in our analysis that must first be defined:

- A *message* is a group of dummy bytes produced by a node's traffic generator for intended transportation to a randomly-selected destination. All of our generated messages are small enough to fit into a single data packet.[6]
- A *data packet* is any transmitted packet containing message data. Therefore, every *message* requires the transmission of at least one *data packet* to reach its destination. A message also can generate more than one data packet, depending on the route length, and the delivery strategy of the algorithm. It also is possible for a message to generate *no* data packets, if the sending node fails to identify any active route toward the message's destination.
- A *control packet* is any transmitted packet that does not contain message data. Control packets are the means by which most algorithms communicate routing information with nearby nodes.

2.3.1 Message delivery ratio

We calculate the message delivery ratio for each algorithm by dividing the total number of messages received at their intended destination by the total number of messages generated. This metric measures each algorithm's overall success in providing reliable communication.

We note that this metric is typically referred to as *packet delivery ratio* in similar examinations of routing protocol behavior. In this study, however, we substitute the word *message* for *packet* to keep our terminology consistent with the precise definitions provided above.

Figure 1 shows the message delivery ratio for each of the four algorithms. A striking result from this comparison is the dominance of ODMRP. This high delivery rate is best explained by ODMRP's aggressive flooding approach to route discovery. Instead of using control packets to discover message routes, this algorithm floods the network with data packets. This greatly increases the chance that a message will reach its intended destination.

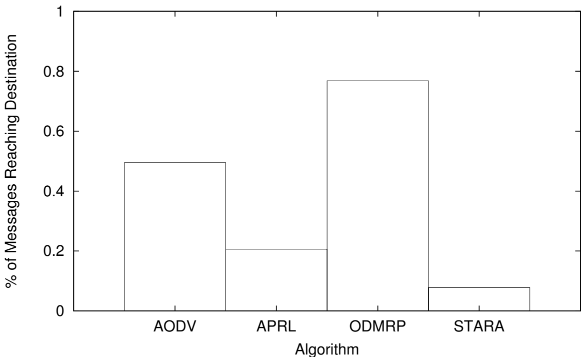


Figure 1: Message delivery ratio comparison of all four algorithms.

We also note that AODV performs better than APRL. This result indicates the advantage of a reactive approach to route discovery in our mobile scenario. If the nodes in our experiment had been more static, or if the physical environment had otherwise provided less opportunity for link breakages, it is possible that we would see less of a gap between AODV's and APRL's delivery ratios. In addition, APRL does not minimize a hop count metric when choosing a route. Subsequently, its average hop count (shown in Section 2.3.3 below) for successful message transmissions is larger than what we see for AODV. This use of longer routes opens up more opportunity for dropped packets, and therefore it also may have lowered APRL's performance.

STARA's message delivery ratio is the worst of the group. We attribute the algorithm's poor performance in our experiment to an excessive amount of control traffic. Its continual probing of the network created overwhelming congestion. Accordingly, we recognize that our implementation of STARA needs additional flow restrictions on the control traffic to constrain the unchecked propagation of control packets. Piyush Gupta proposes one possible solution to this excessive traffic problem by noting that multiple copies of an identical control packet arriving at a common node could be condensed into a fewer number of packets before being rebroadcast [Gup00]. Gupta also notes that "extensive simulation study and protocol development [are] needed to make STARA a viable routing protocol" [Gup00]. We agree, adding that our experience with STARA reinforces the importance of using detailed stochastic simulation to help validate and enhance routing protocols during the design phase.

2.3.2 Communication efficiency

Figure 2 shows, for each algorithm, the average number of data packets transmitted for each generated message. We derive this number by dividing the total number of trans-

[6] Each message was approximately 1200 bytes in size, including all relevant headers.

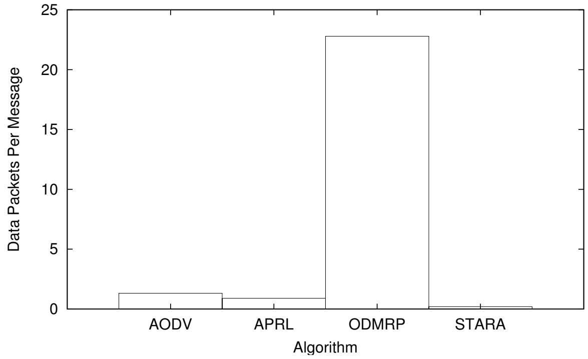


Figure 2: Average number of data packets transmitted for each generated message.

mitted data packets by the total number of generated messages. This metric approximates each algorithm's data transportation efficiency.

ODMRP once again dominates the plot. This is not unexpected, as ODMRP floods the network with data packets when trying to locate a route. Accordingly, the number of data packet transmitted for each message in ODMRP is significant. If the size of the messages being transmitted is large, then the effect of this data packet load on available bandwidth would be dramatic. In our experiment, however, the generated traffic size was relatively modest, approximately 423 outgoing bytes per laptop per second, allowing ODMRP to avoid excessive congestion. If, on the other hand, the traffic had been concentrated on fewer destinations, or if the network had been more static, we would observe fewer ODMRP data packets as it would not need to flood the network as often.

AODV transmitted 1.32 data packets per message, while APRL transmitted only .90. The difference between these two values, though small in magnitude, is notable. As shown below in Section 2.3.3, APRL's average hop count for successfully received messages is larger than AODV. Therefore, if both algorithms had equally accurate routing information, APRL's data packets per message value should be larger than AODV, as its routes tend to require more hops. This is not, however, what we observe. APRL's smaller value of data packets per message indicates a lack of quality routing information. As we explore in more detail in the next section, APRL had a large number of messages dropped at their source (without generating any data packets), because an active route could not be identified.

STARA transmitted the fewest data packets, which we attribute to the packet drops due to the congestion created by the algorithms control traffic.

Figure 3 shows, for each algorithm, the average number of control packets transmitted for each generated message. We derive this number by dividing the total number of transmitted control packets by the total number of generated messages. This allows us to compare each algorithm's control traffic efficiency.

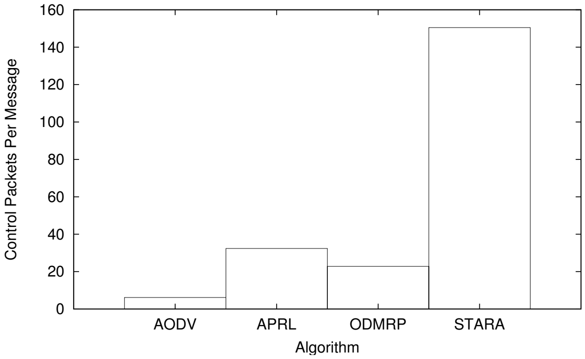


Figure 3: Average number of control packets transmitted for each generated message.

	Packets Per Message
AODV	7.50
APRL	33.30
ODMRP	45.59
STARA	150.67

Table 1: The average number of packets transmitted (data packets and control packets) for each generated message.

STARA clearly produced the most control traffic. It generated, on average, 150 control packets for each message. This overwhelming result evidences the excessive control traffic that we cite as causing STARA's poor performance in our experiment.

APRL generated the next largest amount of control traffic, with 32 control packets, on average, for each message. AODV was the most efficient, generating only six control packets, on average, for each message. These results demonstrate that in our scenario, with light traffic and dynamic connectivity, one of the costs of APRL's periodic proactive route discovery, as oppose to AODV's reactive approach, is a substantial increase in control traffic.

It is difficult to find comparative significance for the ODMRP control traffic result, as control and data packets are not clearly distinguished for this algorithm. In our experiment, we count packets not containing message data as control packets. For ODMRP this would include only the reply traffic generated in response to the algorithm's flooding of the network with data packets, even though, in many ways, the flooded data packets are acting the role of control packets. A more meaningful comparison of ODMRP's communication efficiency can be found with the *total packets per message* values that we present next.

Table 1 shows, for each algorithm, the average number of packets transmitted for each generated message. These values are calculated by adding the total number of control packets and data packets transmitted, and then dividing this sum by the total number of generated messages. This measure approximates the overall communication efficiency of each algorithm.

AODV is clearly the most efficient, requiring, on average, only 7.5 packets for each message. Surprisingly, ODMRP does not fare much worse than APRL. One might assume that ODMRP's aggressive network flooding approach would lead to a more significant increase in traffic costs as compared to APRL's periodic route advertisements. However, ODMRP's 45.59 packets transmitted for each message is not overwhelmingly larger than APRL's 33.30. It should, however, be noted that if the size of the data packets being transmitted is large, APRL would gain a more noticeable lead over ODMRP, as the majority of APRL's traffic is in the form of streamlined control packets,[7] whereas ODMRP includes copies of its data packets with much of its traffic. Considering that AODV and ODMRP are both reactive algorithms that flood the network to discover routes, it is also surprising to note how many fewer packets per message are required by AODV. This result emphasizes that it is important for protocol designers to carefully consider the flow restrictions on their route discovery packets. Finally, we note that this experiment generated a modest amount of messages. Because APRL's control traffic should remain constant regardless of the number of messages being sent, it might gain more of an efficiency advantage over its reactive counterparts if the amount of data traffic was greatly increased.

2.3.3 Hop count

Figure 4 shows, for each algorithm, the average number of hops successfully received messages traveled to reach their destination. We limit our sample to successful messages because we are interested in characterizing the typical route selected by each algorithm.

For ODMRP, it is difficult to calculate hop count values because messages can be received at their destination multiple times from multiple paths of varying length. We avoid this problem in this plot by counting only the first copy of each message to successfully arrive.

STARA has the lowest average hop count for successfully received packets. This result, however, is due to the excessive control traffic congestion which made successful packet transmission difficult. In this environment, only packets being sent to a neighbor had a good chance of succeeding. Therefore we cannot gain a good understanding of the typical route selected by this algorithm in more forgiving conditions.

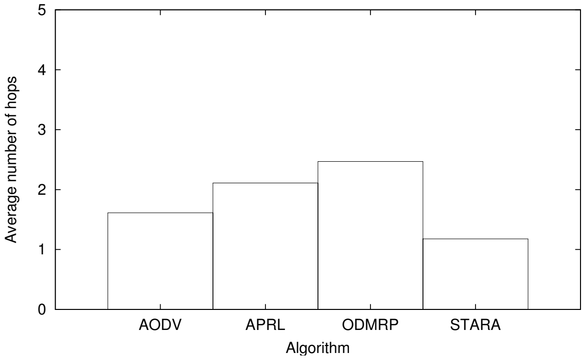


Figure 4: The average number of hops traveled for successfully received messages.

After STARA, AODV required the next fewest number of hops for successful packets. It is expected that AODV should have a lower average hop count than APRL or ODMRP, as the former finds routes on demand, and selects for a shorter path, whereas the latter two do not consider hop count in selecting a route.

In fact, ODMRP has the largest average hop count value. Because many of its messages arrive randomly at their destination during the undirected route discovery phase, ODMRP is unable to always use the most efficient identified route.

Figures 5–7 provide a distribution of hop count values for each algorithm. Specifically, they show the total number of messages that traveled each number of hops. They also include an independent bar for messages that were successfully received, and an independent bar for messages that failed. This allows a more detailed understanding of the relationship between route size and message delivery success. They also include a bar for zero hops, which represent failed message that never left the sending host, for lack of a route. We omit a detailed ODMRP hop count distribution because its flooding approach to message delivery makes it impossible to define a comparable hop count value for failed messages.

Figure 5 shows that the vast majority of STARA messages never made it beyond their source node. Successful messages are clustered almost entirely in the one-hop bucket, with 513 of 598 total successful messages traveling one hop, 73 traveling two, and only 12 successful messages traveling any further. The maximum path length traveled by any successful message was 7 hops. There were also a small number of failed messages that traveled unusually long routes before failing. The longest such route was 33 hops.

Considering that this algorithm was transmitting, on average, 150 packets for each message, interference likely caused the large number of 0-hop message failures that we observe. Interference also limited the ability of the algo-

[7] APRL uses only a simple binary indication of whether or not a route exists in its routing table, leading to small control packets.

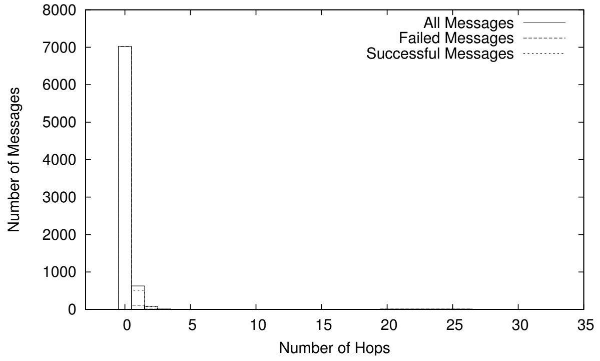


Figure 5: Distribution of STARA hop count values for all messages, successful messages, and failed messages.

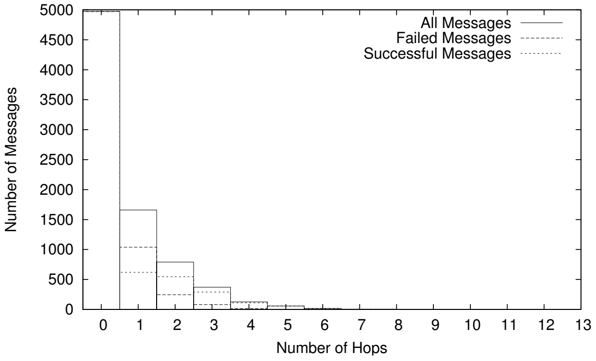


Figure 6: Distribution of APRL hop count values for all messages, successful messages, and failed messages.

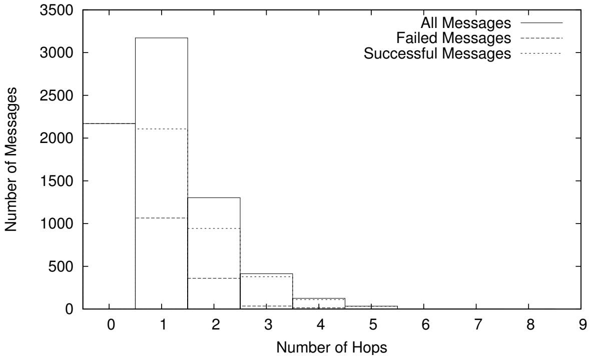


Figure 7: Distribution of AODV hop count values for all messages, successful messages, and failed messages.

rithm to maintain reliable routing information. The failed packets observed to travel over unusually long paths are likely caused by invalid and looping routes. This is not suprising, as the congestion generated by STARA would make it difficult for any node to maintain a full set of consistently valid routes.

Figure 6 shows that APRL also had a significant number of messages fail without leaving their source node. Successful messages are divided almost equally between one and two hops, and failed messages decrease regularly from one to four hops in proportion to the decreasing number of total messages in each of those buckets. The longest path traveled by any message was 12 hops.

The large number of observed 0-hop message failures reveals that the majority of generated messages were dropped because APRL could not identify a valid route to the desired destination. The implication of this striking result is that APRL's periodic route advertising scheme was unable to consistently maintain adequate routing information in our experiment. While we admit that there would be many situations in our experiment where a route physically did not exist between two nodes, the large difference between APRL's and AODV's 0-hop failures indicates this more serious problem (AODV is shown in Figure 7). It is also interesting to note that failed messages outnumber successful messages in the 1-hop bucket. This relationship is the opposite of what we see with AODV. The explanation for this behavior may involve APRL not selecting for shorter routes. With an average route length of 2.11 hops (as compared to AODV's 1.61), APRL creates more opportunity for invalid routes or collisions to create failed packet transmissions after one hop.

Figure 7 shows that AODV has far fewer 0-hop message failures. The majority of its successful messages traveled one hop, but those that traveled two, three, or four hops significantly outnumber the failed messages in their respective buckets. This is especially noticeable in the two and three-hop buckets where almost *all* of the messages which traveled that distance were successful. The longest path traveled by any message was 8 hops.

The implication of these results is that AODV's on-demand approach to route discovery worked well in our dynamic environment. When the algorithm could identify a route, which it did much more frequently than APRL, it was subsequently successful in delivering a message to its intended destination. This indicates quality route information, and the advantage of AODV in finding more one-hop paths than APRL, increasing its overall delivery success, and conserving network resources.

2.3.4 End-to-end latency

Calculating end-to-end latency for ad hoc networks is difficult. The main obstacle is a lack of synchronization between the individual node clocks. This creates a situation in which a comparison of receiver and sender timestamps is not sufficient for generating accurate latency values.

In our experiment we did not run NTP. We made this decision to avoid extra computational overhead and bandwidth usage. Instead, we relied on the GPS units to provide accuracy to our clocks. Specifically, we set each node clock from the GPS units before the experiment, and reset them from the units every 10 seconds during the experiment. Since we required regular GPS queries for other purposes (such as tracking node mobility), this approach did not introduce significant extra overhead, and required no bandwidth usage. We found, however, that our node clocks still drifted from each other on the order of tens to a few hundreds of milliseconds. We attribute this to delays in reading the time from the GPS unit and invoking the kernel time system calls.

Though relatively small, this clock drift is still significant as many of our calculated end-to-end latency values are within the same order of magnitude. In this study, we introduce a novel approach to better approximate time synchronization in mobile ad hoc nodes. We take advantage of the fact that our nodes were configured to broadcast a simple beacon at a regular interval (once every 3 seconds), which provides us with a convenient set of *time synchronization events.* Specifically, if we want to synchronize the clocks between node A and node B at time t, we analyze our beacon logs to find a beacon that was sent by a third node, C, and that was received by both A and B near time t. Assuming that A and B should receive the broadcast beacon more or less at the same instant, we can calculate the skew between the two clocks around time t by comparing what time they each received C's beacon. This concept can be extended to find the clock skew between all node pairs at all times by locating an appropriate time synchronization event for each ($node_A$, $node_B$, $time_t$) 3-tuple.

To be computationally efficient, however, we approximated this calculation by splitting the duration of each algorithm's run into a group of 13 equally-sized time buckets. For each bucket, we calculated the average skew value for every pair of nodes. We did so by performing the skew calculation for every time synchronization event that occurred within the bucket's time range, and then averaging the skew values generated for each pair.

To subsequently calculate the end-to-end latency for a given message sent from A to B, we use the send time to locate the appropriate bucket, and then use the average skew value stored for (A, B) in that bucket to synchronize the clocks. If there are no time synchronization events between A and B during the relevant bucket duration, we throw out the message, and do not include its latency value in our metric.

We recognize that this approach is only approximate for several reasons: 1) it is unrealistic to assume that two nodes A and B will receive and timestamp a broadcast beacon at the same instant, as computational factors unique to each node can affect how long it takes for the event to actually be logged; 2) in a multi-hop routing environment, it is possible that the sender and the receiver of a given packet are too distant to have recently received the same beacon; and 3) using the average observed time skew calculated over a given bucket duration is not as accurate as always searching out the closest single time synchronization event.

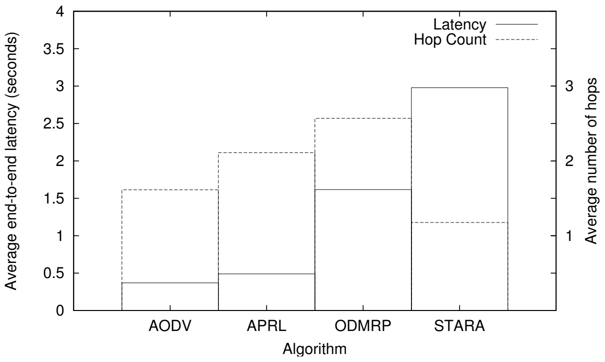


Figure 8: A comparison of average corrected end-to-end latency, plotted with average hop count for successful messages.

Accordingly, we do not present our end-to-end latency values as precise measurements. We do, however, maintain that our corrected values are more accurate indications of transit time than relying on uncorrected timestamps. We use them here to provide meaningful insight on matters such as the relative ranking of the algorithms. We leave the refinement of this technique as future work.

Figure 8 shows the average corrected end-to-end latency value and the average hop count value for successful messages. We find the expected relationship between end-to-end latency and hop count. For AODV, APRL, and ODMRP, the average end-to-end latency value increases roughly proportionately to the average hop count value. STARA is an exception because it shows a low average hop count, and a large average end-to-end latency. This abnormality can be explained by the large amount of computational overhead generated by the excess amount of control traffic. The notable volume of control packets in both the receive and send queues of all nodes could significantly increase the delay between the sender generating a message and the receiver processing it.

2.4 Conclusions

Any conclusions we draw from this outdoor experiment must be qualified by the conditions of our particular testing environment. A markedly different scenario could produce markedly different results. For example, our nodes were highly mobile, and our terrain was non-uniform (though there were few permanent obstructions), leading to a dynamic state of connectivity. This environment may disadvantage an algorithm like APRL that does not seek routes on demand. Similarly, our traffic load was relatively light,[8] which may have advantaged an aggressive data-packet flooding algorithm like ODMRP, which may have failed under heavier traffic conditions.

With these qualifications in mind, we present the following conclusions:

- **AODV is efficient and effective.** Though its message delivery ratio was not as high as ODMRP, it delivered messages significantly better than APRL and STARA. More importantly, on all measures of communication efficiency, AODV generated by far the least amount of traffic for each message. And in terms of route selection, AODV was successful in consistently finding short paths, giving it the additional advantage of having the lowest average end-to-end latency value. In an environment with limited bandwidth, or limited energy resources, AODV is a good choice as a provider of low-cost, adaptable, reliable, and fast communication.

- **ODMRP is optimal for specific scenarios, bad for others.** This algorithm generates a lot of overhead traffic. Its network flooding is bandwidth intensive, and if data packets are large, ODMRP could fail due to congestion. At the same time, however, it had the highest message delivery ratio of all four algorithms. This indicates that in a situation in which bandwidth and energy resources are plentiful, data packets are small, and communication reliability is crucial, ODMRP is a good choice.[9]

- **APRL performed poorly in our environment.** Its message delivery ratio was low, its overhead was large, and it had a substantial percentage of packets fail at their source. Our results indicate that APRL had a hard time maintaining reliable routing information in our relatively dynamic environment. In any scenario comparable to our experiment, APRL shows no clear advantage over a reactive algorithm such as AODV.

- **Our STARA implementation emphasizes the importance of flow control.** In their original paper, Gupta and Kumar validated STARA with a simple stochastic simulation that did not model collision or interference effects [GK97]. Their analytic validation demonstrated that STARA performs better than other approaches because of its dynamic avoidance of highly trafficked routes. Their analysis, however, avoids the reality that would be clear in more detailed simulation: If control traffic is not carefully controlled, it can destabilize the entire network through excessive congestion. Gupta later identifies the potential for this problem in his PhD thesis, where he briefly suggests one possible solution [Gup00]. He goes on to suggest that more extensive simulation is necessary before the design could be considered complete. We agree, and further recommend that *all* protocol designers integrate more detailed simulation into their design process so as to more effectively address necessary practical concerns, such as flow control, in their original protocol specifications.

- **Reactive is better than proactive in dynamic environments.** APRL and STARA's poor performance, as compared to AODV and ODMRP's relative success, highlights the general advantage of a reactive approach to routing in a dynamic environment. Our analysis of APRL shows an unnecessarily large number of messages dropped before leaving their source node, and STARA crippled itself with excessive proactive discovery. It is a fair assumption that if we had restrained STARA's control traffic to a reasonable level, it would have faced the same lack of quality routing information demonstrated by APRL. Similarly, if we had decreased APRL's route advertisement interval to increase the timeliness of its routing information, it would have suffered from an excess amount of control traffic. This observation underscores the perhaps unresolvable tension between control traffic and message delivery success present in proactive algorithms operating in dynamic environments: If you make your algorithm efficient, its reliability drops; if you make your algorithm reliable, its efficiency drops. Reactive approaches are clearly preferable for scenarios with variable connectivity.

3 Common assumptions in ad hoc network simulation

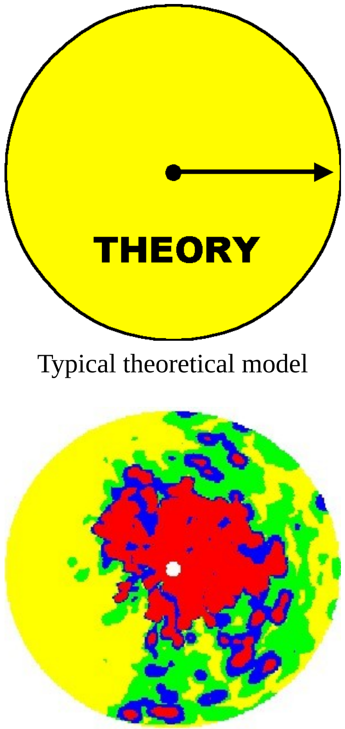
Now that we have described the behavior of a real ad hoc network, we can explore how close simulation comes to reproducing this reality. In this section, we demonstrate how the commonly used theoretical models of radio behavior are far simpler than reality, and we codify these simplifications into a group of six assumptions commonly made by simulation designers. Using data from our outdoor experiment, we prove these six assumptions to be false.

In the following section, we use a wireless network simulator, configured to mimic our real world experiment,

[8] Messages were generated, on average, only once every three seconds.

[9] Because we reduced ODMRP to the unicast case for our experiment, we can not specifically address its effectiveness as a provider of multicast communication. We hope, however, that our analysis of its general communication efficiency and reliability can still act as useful guides for those interested in effective multicast communication.

to quantify and describe the impact of these assumptions on simulation results.



Typical theoretical model

Source: Comgate Engineering
http://www.comgate.com/ntdsign/wireless.html

Figure 9: Real radios, such as the one at the bottom, are more complex than the common theoretical model at the top. Here different colors, or shades of gray, represent different signal qualities.

3.1 Radios in theory and practice

The top example in Figure 9 provides a simple model of radio propagation, one that is used in many simulations of ad hoc networks; contrast it to the bottom example of a real signal-propagation map, drawn at random from the web. Measurements from an ad hoc network of Berkeley Motes demonstrate a similar non-uniform non-circular behavior [GKW$^+$02]. The simple model is based on Cartesian distance in an X-Y plane. More realistic models take into account antenna height and orientation, terrain and obstacles, surface reflection and absorption, and so forth.

Of course, not every simulation study needs to use the most detailed radio model available, nor explore every variation in the wide parameter space afforded by a complex model. The level of detail necessary for a given analytic or simulation study depends on the characteristics of the study. The majority of results published to date use the simple models, however, with no examination of the sensitivity of results to the (often implicit) assumptions embedded in the model.

There are real risks to protocol designs based on overly simple models of radio propagation. First, “typical” network connectivity graphs look quite different in reality than they do on a Cartesian grid. An antenna placed top of a hill has direct connectivity with all other nearby radios, for example, an effect that cannot be observed in simulations that represent only flat plains. Second, it is often difficult in reality to estimate whether or not one has a functioning radio link between nodes, because signals fluctuate greatly due to mobility and fading as well as interference. Broadcasts are particularly hard-hit by this phenomenon as they are not acknowledged in typical radio systems. Protocols that rely on broadcasts (e.g., beacons) or “snooping” may therefore work significantly worse in reality than they do in simulation.

Figure 10 depicts one immediate drawback to the over-simplified model of radio propagation. The three different models in the figure, the Cartesian “Flat Earth” model, a three-dimensional model that includes a single hill, and a model that includes (absorptive) obstacles, all produce entirely different connectivity graphs, even though the nodes are in the same two-dimensional positions. As all the nodes move, the ways in which the connectivity graph changes over time will be different in each scenario.

Figure 11 presents a further level of detail. At the top, we see a node’s trajectory past the theoretical (T) and practical (P) radio range of another node. Beneath it we sketch the kind of change in link quality we might expect under these two models. The theoretical model (T) gives a simple step function in connectivity: either one is connected or one is not. Given a long enough straight segment in a trajectory, this leads to a low rate of change in link connectivity. As such, this model makes it easy to determine when two nodes are, or are not, “neighbors” in the ad hoc network sense.

In the more realistic model (P), the quality of the link is likely to vary rapidly and unpredictably, even when two radios are nominally “in range.” In these more realistic cases, it is by no means easy to determine when two nodes have become neighbors, or when a link between two nodes is no longer usable and should be torn down. In the figure, suppose that a link quality of 50% or better is sufficient to consider the nodes to be neighbors. In the diagram, the practical model would lead to the nodes being neighbors briefly, then dropping the link, then being neighbors again, then dropping the link.

In addition to spatial variations in signal quality, a radio’s signal quality varies over time, even for a stationary radio and receiver. Obstacles come and go: people and vehicles move about, leaves flutter, doors shut. Link connectivity can come and go; one packet may reach a neighbor successfully, and the next packet may fail. Both

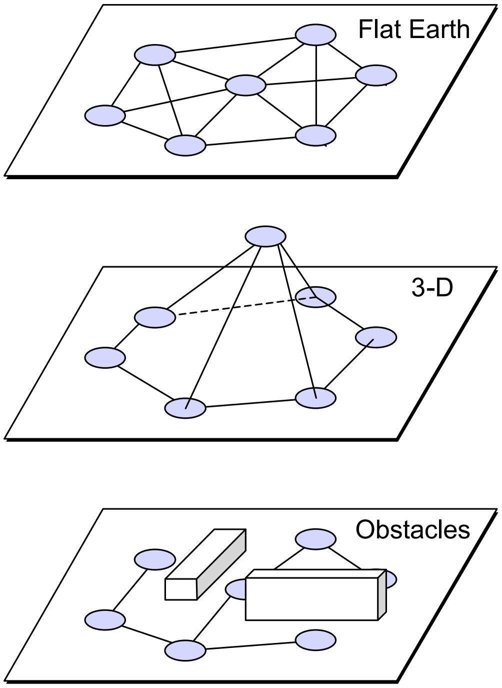


Figure 10: The Flat Earth model is overly simplistic.

short-term and long-term changes are common in reality, but not considered by most practical models. Some, but not all, of this variation can be masked by the physical or data-link layer of the network interface.

Although the theoretical model may be easy to use when simulating ad hoc networks, it leads to an incorrect sense of the way the network evolves over time. For example, in Figure 11, the link quality (and link connectivity) varies much more rapidly in practice than in theory. Many algorithms and protocols may perform much more poorly under such dynamic conditions. In some, particularly if network connectivity changes rapidly with respect to the distributed progress of network-layer or application-layer protocols, the algorithm may fail due to race conditions or a failure to converge. Simple radio models fail to explore these critical realities that can dramatically affect performance and correctness. For example, Ganesan et al. measured a dense ad hoc network of sensor nodes and found that small differences in the radios, the propagation distances, and the timing of collisions can significantly alter the behavior of even the simplest flood-oriented network protocols [GKW+02].

In summary,“good enough” radio models are quite important in simulation of ad hoc networks. The Flat Earth model, however, is by no means good enough. In the following sections we make this argument more precise.

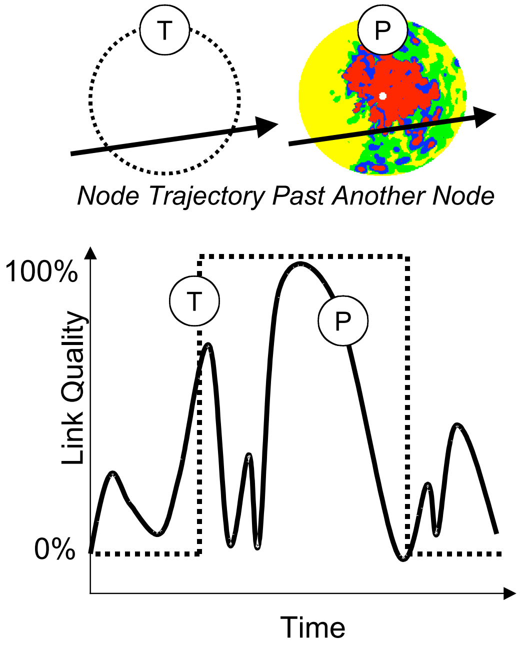


Figure 11: Difference between theory (T) and practice (P).

3.2 Models used in research

We surveyed a set of MobiCom and MobiHoc proceedings from 1995 through 2003. We inspected the simulation sections of every article in which RF modeling issues seemed relevant, and categorized the approach into one of three bins: *Flat Earth*, *Simple*, and *Good*. This categorization required a fair amount of value judgment on our part, and we omitted cases in which we could not determine these basic facts about the simulation runs.

Figure 12 presents the results. Note that even in the best years, the Simple and Flat-Earth papers significantly outnumber the Good papers. A few, such as Takai et al. [TMB01], deserve commendation for thoughtful channel models.

Flat Earth models are based on Cartesian X–Y proximity, that is, nodes A and B communicate if and only if node A is within some distance of node B.

Simple models are, almost without exception, `ns-2` models using the CMU 802.11 radio model [FV02]. This model provides what has sometimes been termed a “realistic” radio propagation model. Indeed it is significantly more realistic than the “Flat Earth” model, e.g., it models packet delay and loss caused by interference rather than assuming that all transmissions in range are received perfectly. We still call it a “simple” model, however, because it embodies many of the questionable axioms we detail

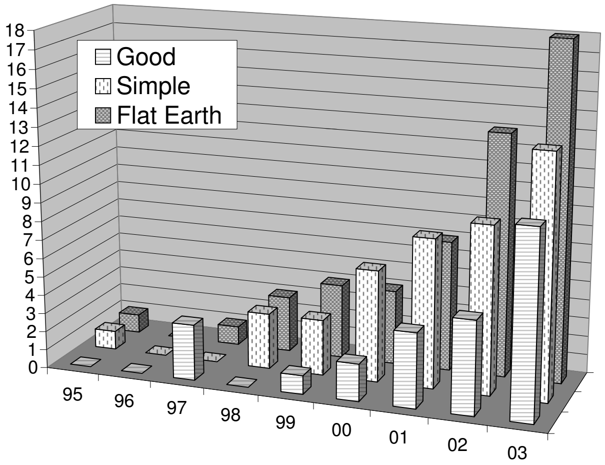


Figure 12: The number of papers in each year of Mobicom and MobiHoc that fall into each category.

below. In particular, the standard release of ns-2 provides a simple free-space model (often termed a "Friis-free-space" model in the literature) and a two-ray ground-reflection model. Both are described in the ns-2 document package [FV02].

The free-space model is similar to the "Flat Earth" model described above, as it does not include effects of terrain, obstacles, or fading. It does, however, model signal strength with somewhat finer detail ($1/r^2$) than just "present" or "absent."

The two-ray ground-reflection model, which considers both the direct and ground-reflected propagation path between transmitter and receiver, is better, but not particularly well suited to most MANET simulations. It has been reasonably accurate for predicting large-scale signal strength over distances of several kilometers for cellular telephony systems using tall towers (heights above 50m), and also for line-of-sight micro-cell channels in urban environments. Neither is characteristic of typical MANET scenarios. In addition, while this propagation model does take into account antenna heights of the two nodes, it assumes that the earth is flat (and there are otherwise no obstructions) between the nodes. This may be a plausible simplification when modeling cell towers, but not when modeling vehicular or handheld nodes because these are often surrounded by obstructions. Thus it too is a "Flat Earth" model, even more so if the modeler does not explicitly choose differing antenna heights as a node moves.[10]

More recently, ns-2 added a third channel model—the "shadowing" model described earlier by Lee [Lee82]—to account for indoor obstructions and outdoor shadowing via a probabilistic model [FV02]. The problem with ns-2's shadowing model is that the model does not consider correlations: a real shadowing effect has strong correlations between two locations that are close to each other. More precisely, the shadow fading should be modeled as a two-dimensional log-normal random process with exponentially decaying spatial correlations (see [Gud91] for details). To our knowledge, only a few simulation studies include a valid shadowing model. For example, WiPPET considers using the correlated shadowing model to compute a gain matrix to describe radio propagation scenarios [KLM$^+$00]. WiPPET, however, only simulates cellular systems. The simulation model we later use for this study considers the shadowing effect as a random process that is temporally correlated; between each pair of nodes we use the same sample from the log-normal distribution if the two packets are transmitted within a pre-specified time period.[11]

Good models have fairly plausible RF propagation treatment. In general, these models are used in papers coming from the cellular telephone community, and concentrate on the exact mechanics of RF propagation. To give a flavor of these "good" models, witness this quote from one such paper [ER00]:

> In our simulations, we use a model for the path loss in the channel developed by Erceg et al. This model was developed based on extensive experimental data collected in a large number of existing macro-cells in several suburban areas in New Jersey and around Seattle, Chicago, Atlanta, and Dallas.... [Equation follows with parameters for antenna location in 3-D, wavelength, and six experimentally determined parameters based on terrain and foliage types.] ...In the results presented in this section, ...the terrain was assumed to be either hilly with light tree density or flat with moderate-to-heavy tree density. [Detailed parameter values follow.]

Of course, the details of RF propagation are not always essential in good network simulations; most critical is the overall realism of connectivity and changes in connectivity (Are there hills? Are there walls?). Along these lines, we particularly liked the simulations of well-known routing algorithms presented by Johansson et al. [JLH$^+$99], which used relatively detailed, realistic scenarios for a conference room, event coverage, and disaster area. Although this paper employed the ns-2 802.11 radio model, it was rounded out with realistic network obstacles and node mobility.

[10] See also Lundberg [Lun02], Sections 4.3.4–4.3.5, for additional remarks on the two-ray model's lack of realism.

[11] A recent study by Yuen et al. proposes a novel approach to modeling the correlation as a Gauss-Markov process [YLA02]. We are currently investigating this approach.

3.3 Common MANET axioms

For the sake of clarity, let us be explicit about some basic "axioms" upon which most MANET research explicitly or implicitly relies. These axioms, not all of which are orthogonal, deeply shape how network protocols behave. We note that all of these axioms are contradicted by the actual measurements reported in the next section.

0: The world is flat.
1: A radio's transmission area is circular.
2: All radios have equal range.
3: If I can hear you, you can hear me (symmetry).
4: If I can hear you at all, I can hear you perfectly.
5: Signal strength is a simple function of distance.

There are many combinations of these axioms seen in the literature. In extreme cases, the combination of these axioms leads to a simple model like that in the top diagram in Figure 9. Some papers assume Axioms 0–4 and yet use a simple signal propagation model that expresses some fading with distance; a threshold on signal strength determines reception. Some papers assume Axioms 0–3 and add a reception probability to avoid Axiom 4.

In this paper we address the research community interested in ad hoc routing protocols and other distributed protocols at the network layer. The network layer rests on the physical and medium-access (MAC) layers, and its behavior is strongly influenced by their behavior. Indeed many MANET research projects consider the physical and medium-access layer as a single abstraction, and use the above axioms to model their combined behavior. We take this network-layer point of view through the remainder of the section. Although we mention some of the individual physical- and MAC-layer effects that influence the behavior seen at the network layer, we do not attempt to identify precisely which effects cause which behaviors; such an exercise is beyond the scope of this paper. We next show that the above axioms do not adequately describe the network-layer's view of the world. Then, in Section 4, we show how the use of these axioms leads simulations to results that differ radically from reality.

3.4 The reality of the axioms

Unfortunately, real wireless network devices are not nearly as simple as those considered by the axioms in the preceding section. In this section, we use data collected from the large MANET experiment described previously to examine the reality of radio behavior in an actual ad hoc network implementation. We demonstrate how this reality clearly differs from the behavior described by our axioms.

Before proceeding, it should be noted that the wireless cards in our experiment operated at 2 Mb/s. This fixed rate made it much easier to conduct the experiment, since we did not need to track (and later model) automatic changes to each card's transmission rate. Most current wireless cards are multi-rate, however, which could lead to **Axiom 6: Each packet is transmitted at the same bit rate.** We leave the effects of this axiom as an area for future work.

We also note that in the following analysis we do not use data from the STARA portion of the outdoor experiment. We were concerned that the excessive control traffic generated by this algorithm might impede an accurate assessment of the observed radio behavior.

3.4.1 Axiom 0

The world is flat.

Common stochastic radio propagation models assume a flat earth, and yet clearly the Earth is not flat. Even at the short distances considered by most MANET research, hills and buildings present obstacles that dramatically affect wireless signal propagation. Furthermore, the wireless nodes themselves are not always at ground level. A local researcher using Berkeley "motes" for sensor-network research notes the critical impact of elevation and ground-reflection effects:

> In our current experiments we just bought 60 plastic flower pots to raise the motes off the ground because we found that putting the motes on the ground drastically reduces their transmit range (though not the receive range). Raising them a few inches makes a big difference.

Even where the ground is nearly flat, note that wireless nodes are often used in multi-story buildings. Indeed two nodes may be found at exactly the same x, y location, but on different floors. (This condition is common among the WiFi access points deployed on our campus.) Any Flat Earth model would assume that they are in the same location, and yet they are not. In some tall buildings, we found it was impossible for a node on the fourth floor to hear a node in the basement, at the same x, y location.

We need no data to "disprove" this axiom. Ultimately, it is the burden of all MANET researchers to either a) use a detailed and realistic terrain model, accounting for the effects of terrain, or b) clearly condition their conclusions as being valid only on flat, obstacle-free terrain.

3.4.2 Axioms 1 and 2

A radio's transmission area is circular.

All radios have equal range.

The real-world radio map of Figure 9 makes it clear that the signal coverage area of a radio is far from simple.

Not only is it neither circular nor convex, it often is non-contiguous.

We combine the above two intuitive axioms into a more precise, testable axiom that corresponds to the way the axiom often appears (implicitly) in MANET research.

Testable Axiom 1. The success of a transmission from one radio to another depends only on the distance between radios.

Although it is true that successful communication usually becomes less likely with increasing distance, there are many other factors: (1) All radios are not identical. Although in our experiment we used "identical" WiFi cards, there are reasonable applications where the radios or antennas vary from node to node. (2) Antennas are not perfectly omnidirectional. Thus, the angle of the sender's antenna, the angle of the receiver's antenna, and their relative locations all matter. (3) Background noise varies with time and location. Finally, (4) there are hills and obstacles, including people, that block or reflect wireless signals (that is, Axiom 0 is false).

From the point of view of the network layer, these physical-layer effects are compounded by MAC-layer effects, notably, that collisions due to transmissions from other nodes in the ad hoc network (or from third parties outside the set of nodes forming the network) reduce the transmission success in ways that are unrelated to distance. In this section, we use our experimental data to examine the effect of antenna angle, sender location, and sender identity on the probability distribution of beacon reception over distance.

We first demonstrate that the probability of a beacon packet being received by nearby nodes depends strongly on the angle between sender and receiver antennas. In our experiments, we had each student carry their "node," a closed laptop, under their arm with the wireless interface (an 802.11b device in PC-card format) sticking out in front of them. By examining successive location observations for the node, we compute the orientation of the antenna (wireless card) at the time it sent or received a beacon. Then, we compute two angles for each beacon: the angle between the sender's antenna and the receiver's location, and the angle between the receiver's antenna and the sender's location. Figure 13 illustrates the first of these two angles, while the second is the same figure except with the labels Source and Destination transposed. Figure 14 shows how the beacon-reception probability varied with both angles.

To compute Figure 14, we consider all possible values of each of the two angles, each varying from $[-180, 180)$. We divide each range into buckets of 45 degrees, such that bucket 0 represents angles in $[0, 45)$, bucket 45 represents angles in $[45, 90)$, and so forth. Since we bucket both angles, we obtain the two-dimensional set of buckets shown in the figure. We use two counters for each bucket, one accounting for actual receptions, and the other for potential receptions (which includes actual receptions). Each time a node sends a beacon, every other laptop is a potential recipient. For every other laptop, therefore, we add one to the potential-reception count for the bucket representing the angles between the sender and the potential recipient. If we can find a received beacon in the potential recipient's beacon log that matches the transmitted beacon, we also add one to the actual-reception count for the appropriate count. The beacon reception ratio for a bucket is thus the number of actual receptions divided by the number of potential receptions. Each beacon-reception probability is calculated without regard to distance, and thus represents the reception probability across all distances. In addition, for all of our axiom analyses, we considered only the western half of the field, and incremented the counts only when both the sender and the (potential) recipient were in the western half. By considering only the western half, which is perfectly flat and does not include the lower-altitude section, we eliminate the most obvious terrain effects from our results. Overall, there were 40,894 beacons transmitted in the western half of the field, and after matching and filtering, we had 275,176 laptop pairs, in 121,250 of which the beacon was received, and in 153,926 of which the beacon was not received.

Figure 14 shows that the orientation of both antennas was a significant factor in beacon reception. Of course, there is a direct relationship between the antenna angles and whether the sender or receiver (human or laptop) is between the two antennas. With a sender angle of 180, for example, the receiver is directly behind the sender, and both the sender's body and laptop serves as an obstruction to the signal. A different kind of antenna, extending above the level of the participants' heads, would be needed to separate the angle effects into two categories, effects due to human or laptop obstruction, and effects due to the irregularity of the radio coverage area.

Although the western half of our test field was flat, we observed that the beacon-reception probability distribution varied in different areas. We subdivided the western half into four equal-sized quadrants (northwest, northeast, southeast, southwest), and computed a separate reception probability distribution for beacons sent from each quadrant. Figure 15 shows that the distribution of beacon-reception probability was different for each quadrant, by about 10–15 percent for each distance. We bucketed the laptop pairs according to the distance between the sender and the (intended) destination—the leftmost bar in the graph, for example, is the reception probability for laptop pairs whose separation was in the range $[0, 25)$. Although there are many possible explanations for this quadrant-based variation, whether physical terrain, external noise, or time-varying conditions, the difference between distri-

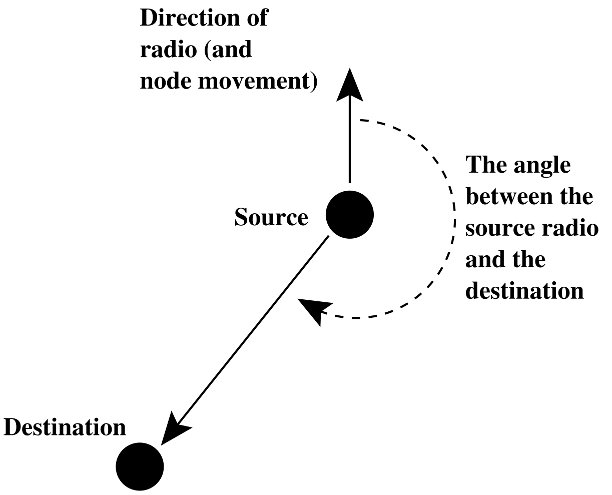

Direction of radio (and node movement)
The angle between the source radio and the destination
Source
Destination

Figure 13: The angle between the sending laptop's antenna (wireless card) and the destination laptop.

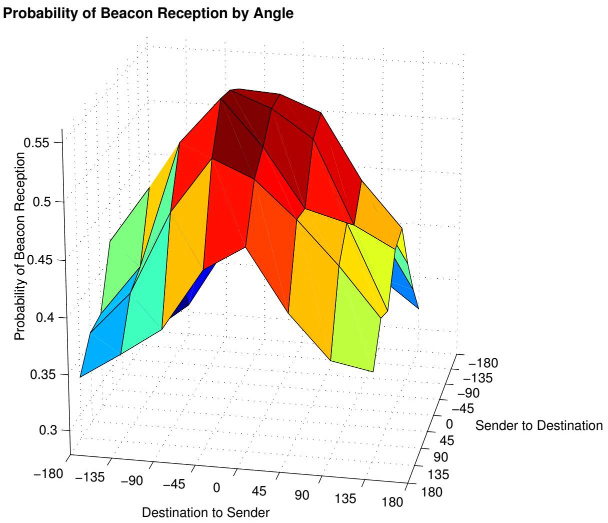


Figure 14: The probability of beacon reception (over all distances) as a function of the two angles, the angle between the sender's antenna orientation and the receiver's location, and the angle between the receiver's antenna orientation and the sender's location. In this plot, we divide the angles into buckets of 45 degrees each, and include only data from the western half of the field. We also express the angles on the scale of -180 to 180, rather than 0 to 360, to better capture the inherent symmetry. -180 and 180 both refer to the case where the sending antenna is pointing directly away from the intended destination, or, correspondingly, the receiving antenna is pointing directly away from the sending node.

butions is enough to make it clear that the location of the sender is not to be ignored.

The beacon-reception probability in the western half of the field also varied according to the identity of the sender. Although all equipment used in every node was an identical model purchased in the same lot and configured identically, the distribution was different for each sender. Figure 16 shows the mean and standard deviation of beacon-reception probability computed across all sending nodes, for each bucket between 0 and 300 meters. The buckets between 250 and 300 meters were nearly empty. Although the mean across nodes, depicted by the boxes, is steadily decreasing, there also is substantial variation across nodes, depicted by the standard-deviation bars on each bucket. This variation cannot be explained entirely by manufacturing variations within the antennas, and likely includes terrain, noise and other factors, even on our space of flat, open ground. It also is important to note, however, that there are only 500-1000 data points for each (laptop, destination bucket) pair. With this number of data points, the differences may not be statistically significant. In particular, if a laptop is moving away from most other laptops, we might cover only a small portion of the possible angles, leading to markedly different results than for other laptops. Overall, the effect of identity on transmission behavior bears further study with experiments specifically designed to test it.

In other work, Ganesan et al. used a network of Berkeley "motes" to measure signal strength of a mote's radio throughout a mesh of mote nodes [GKW+02].[12] The resulting contour map is not circular, nor convex, nor even monotonically decreasing with distance. Indeed, since the coverage area of a radio is not circular, it is difficult to even define the "range" of a radio.

3.4.3 Axiom 3

If I can hear you, you can hear me (symmetry).

More precisely,

Testable Axiom 3: If a message from A to B succeeds, an immediate reply from B to A succeeds.

This wording adds a sense of time, since it is clearly impossible (in most MANET technologies) for A and B to transmit at the same time and result in a successful message, and since A and B may be moving, it is important to consider symmetry over a brief time period so that A and B have not moved apart.

There are many factors affecting symmetry, from the point of view of the network layer, including the physical effects mentioned above (terrain, obstacles, relative antenna angles) as well as MAC-layer collisions. It is

[12] The Berkeley mote is currently the most common research platform for real experiments with ad hoc sensor networks.

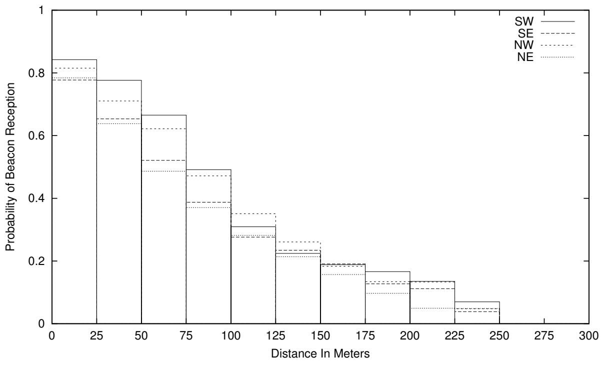


Figure 15: The probability of beacon reception varied from quadrant to quadrant within the western half of the field.

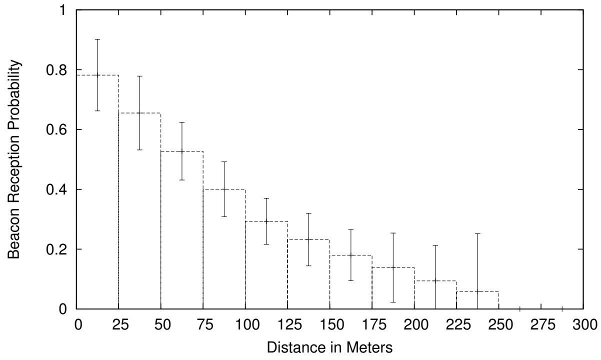


Figure 16: The average and standard deviation of reception probability across all nodes, again for the western half of the field.

worth noting that the 802.11 MAC layer includes an internal acknowledgment feature, and a limited amount of retransmission attempts until successful acknowledgment. Thus, the network layer does not perceive a frame as successfully delivered unless symmetric reception was possible. Thus, for the purposes of this axiom, we chose to examine the broadcast beacons from our experimental dataset, since the 802.11 MAC has no internal acknowledgment for broadcast frames. Since all of our nodes sent a beacon every three seconds, we were able to identify symmetry as follows: whenever a node B received a beacon from node A, we checked to see whether B's next beacon was also received by node A.

Figure 17 shows the conditional probability of symmetric beacon reception. Using the definition of symmetry described above, we calculate each bar by dividing the number of observed symmetric relationships by the total number of observed symmetric *and* asymmetric relationships for the given distance range. If the physical and MAC layer behavior was truly symmetric, this probability would be 1.0 across all distances. In reality, the probability was never much more than 0.8, most likely due to MAC-layer collisions between beacons. Since this graph depends on the joint probability of a beacon arriving from A to B and then another from B to A, the lower reception probability of higher distances leads to a lower joint probability and a lower conditional probability. The abnormal bump in the 200 to 225 meter distance bucket is explained by the fact that the experimental field was roughly 225 meters long on its north-south axis. We observed that it was a common movement pattern to walk to the either the northern or southern terminus of the field, and then turn to head toward another location. Therefore, there commonly occurred a situation where two nodes would be facing each other from opposite ends of the field. In this orientation their reception probability was increased, bumping up the overall probability observed for this range.

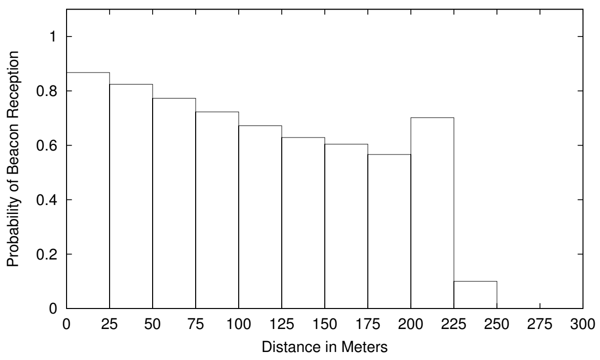


Figure 17: The conditional probability of symmetric beacon reception as it varied with the distance between two nodes, again for the western half of the field.

Figure 18 shows how the conditional probability varied across all the nodes in the experiment. The probability was consistently close to its mean 0.76, but did vary from node to node with a standard deviation of 0.029 (or 3.9%). Similarly, when calculated for each of the four quadrants (not shown), the probability also was consistently close to its mean 0.76, but did have a standard deviation of 0.033 (or 4.3%). As mentioned in the discussion of Axioms 1 and 2, there are many possible explanations for these variations, including physical terrain, external noise, and different movement patterns. Regardless of the specific causes, the fact that this variation exists evidences the invalidity of assuming equal symmetry among all nodes and locations in a real environment.

In other work, Ganesan et al. [GKW$^+$02] noted that about 5–15% of the links in their ad hoc sensor network were asymmetric. In that paper, an asymmetric link had a "good" link in one direction (with high probability of message reception) and a "bad" link in the other direction (with a low probability of message reception). [They do not have a name for a link with a "mediocre" link in either

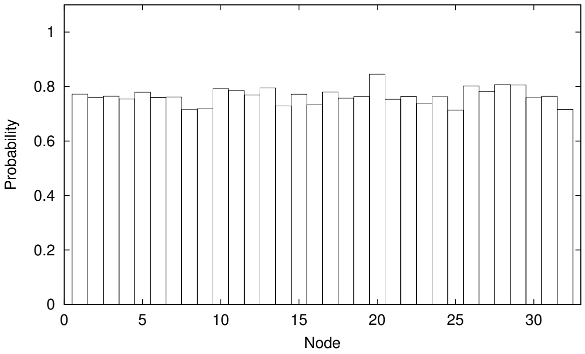


Figure 18: The conditional probability of symmetric beacon reception as it varied across individual nodes, again for the western half of the field.

direction.]

Overall, it is clear that reception is far from symmetric. Nonetheless, many researchers assume this axiom is true, and that all network links are bidirectional. Some do acknowledge that real links may be unidirectional, and usually discard those links so that the resulting network has only bidirectional links. In a network with mobile nodes or in a dynamic environment, however, link quality can vary frequently and rapidly, so a bidirectional link may become unidirectional at any time. It is best to develop protocols that do not assume symmetry.

3.4.4 Axiom 4

If I can hear you at all, I can hear you perfectly.

Testable Axiom 4: The reception probability distribution over distance exhibits a sharp cliff; that is, under some threshold distance (the "range") the reception probability is 1 and beyond that threshold the reception probability is 0.

Looking back at Figure 16, we see that the beacon-reception probability does indeed fade with the distance between the sender and the receiver, rather than remaining near 1 out to some clearly defined "range" and then dropping to zero. There is no visible "cliff." The common `ns-2` model, however, assumes that frame transmission is perfect, within the range of a radio, and as long as there are no collisions. Although `ns-2` provides hooks to add a bit-error-rate (BER) model, these hooks are unused. More sophisticated models do exist, particularly those developed by Qualnet and the GloMoSim project[13] that are being used to explore how sophisticated channel models affect simulation outcomes.

Takai examines the effect of channel models on simulation outcomes [TBTG01], and also concluded that different physical layer models can have dramatically different effect on the simulated performance of protocols [TMB01], but lack of data prevented them from further validating simulation results against real-world experiment results, which they left as future work. In the next section, we compare the simulation results with data collected from a real-world experiment, and recommend that simple models of radio propagation should be avoided whenever comparing or verifying protocols, unless that model is known to specifically reflect the target environment.

3.4.5 Axiom 5

Signal strength is a simple function of distance.

Rappaport [Rap96] notes that the average signal strength should fade with distance according to a power-law model. While this is true, one should not underestimate the variations in a real environment caused by obstruction, reflection, refraction, and scattering. In this section, we show that there is significant variation for individual transmissions.

Testable Axiom 5: We can find a good fit between a simple function and a set of (distance, signal strength) observations.

To examine this axiom, we consider only received beacons, and use the recipient's signal log to obtain the signal strength associated with that beacon. More specifically, the signal log actually contains per-second entries, where each entry contains the single strength of the most recent packet received from each laptop. If a data or routing packet arrives immediately after a beacon, the signal-log entry actually will contain the signal strength of that second packet. We do not check for this situation, since the signal information for the second packet is just as valid as the signal information for the beacon. It is best, however, to view our signal values as those observed within one second of beacon transmission, rather than the values associated with the beacons themselves.

As a starting point, Figure 19 shows the *mean* beacon signal strength observed during the experiment as a function of distance, as well as best-fit linear and power curves. The power curve is a good fit and validates Rappaport's observation. When we turn our attention to the signal strength of individual beacons, however, as shown in Figures 20 and 21, there clearly is no simple (non-probabilistic) function that will adequately predict the signal strength of an individual beacon based on distance alone.

The reason for this difficulty is clear: our environment, although simple, is full of obstacles and other terrain features that attenuate or reflect the signal, and the cards

[13] http://www.scalable-networks.com/pdf/mobihocpreso.pdf

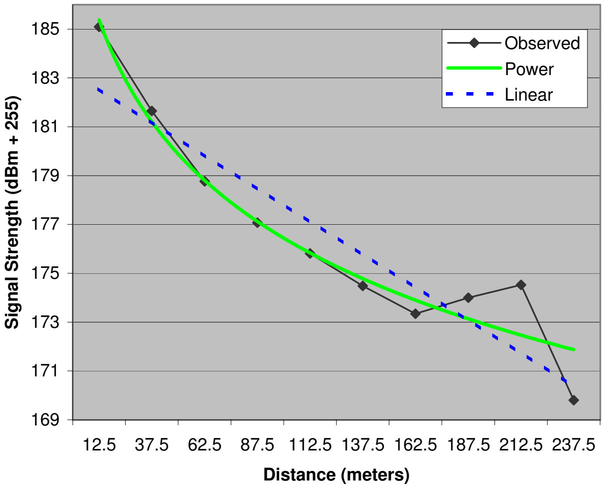


Figure 19: Linear and power-curve fits for the mean signal strength observed in the western half of the field. Note that we show the signal strength as reported by our wireless cards (which is dBm scaled to a positive range by adding 255), and we plot the mean value for each distance bucket at the midpoint of that bucket.

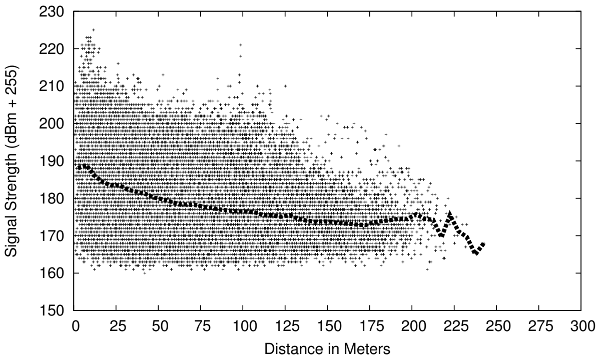


Figure 20: A scatter plot demonstrating the poor correlation between signal strength and distance. We restrict the plot to beacons both sent and received on the western half of the field, and show the mean signal strength as a heavy dotted line.

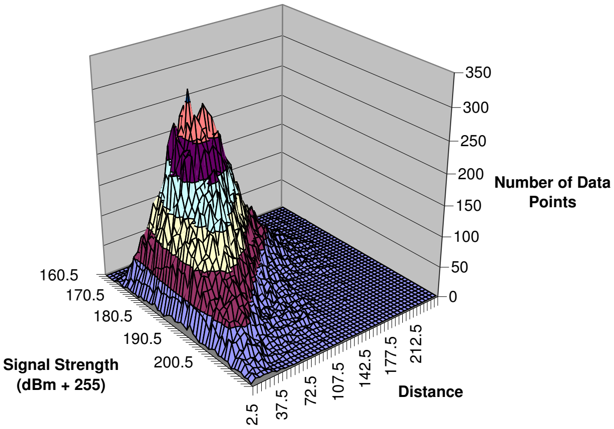


Figure 21: Same as Figure 20 except that it shows the *number* of observed data points as a function of distance and signal strength. There is significant weight relatively far away from the mean value.

themselves do not necessarily radiate with equal power in all directions. In our case, the most common obstacles were the people and laptops themselves, and in fact, we initially expected to discover that the signal strength was better behaved across a specific angle range (per Figure 14) than across all angles. Even for the seemingly good case of both source and destination angles between 0 and 45 degrees (i.e., the sender and receiver roughly facing each other), we obtain a distribution (not shown) remarkably similar to Figure 20. Other angle ranges also show the same distribution as Figure 20.

Overall, noise-free, reflection-free, obstruction-free, uniformly-radiating environments are simply not real, and signal strength of individual transmissions will never be a simple function of distance. Researchers must be careful to consider how sensitive their simulation results are to signal variations, since their algorithms will encounter significant variation once deployed.

Summary. These axioms are often considered to be a reasonable estimate of how radios actually behave, and therefore they are frequently used in simulation without reservation. Our data, however, reveal the danger of such a belief. These assumptions do not just simplify reality, in many cases they distort it. An algorithm that performs well in the calm and predictable physical environment described by our axioms may perform quite differently in the inconsistent and highly variable physical environment that we observe in the real world. These results should compel simulation designers to carefully consider and condition the simplifications they integrate into their radio models. In the next section we quantitatively explore the impact of these axioms.

4 Computer simulation results

We demonstrate above that the axioms are untrue, but a key question remains: what is the effect of these axioms on the quality of simulation results? In this section, we begin by comparing the results of our outdoor experiment with the results of a best-effort simulation model, and then progressively weaken the model by assuming some of the axioms. To better understand the observed effects, we then use a connectivity trace, derived from the outdoor experiment data, to more generally validate and probe the predictive power of our simplified models. We also examine the role of two important parameters in generating the results.

With the exception of the experiment represented by Figure 26, we exclude STARA from our simulation runs because of its excessive control traffic problems. This unusual behavior makes it a poor choice for most of our real world versus simulation comparisons. We include it in this figure alone to validate our claim from Section 2 that detailed simulation would have revealed STARA's traffic flow problems to its designers. As we expected, the detailed model came much closer than the two simpler models in predicting the algorithm's poor performance.[14]

The purpose of this study is not to claim that our simulator can accurately model the real network environment. Instead, we show quantitatively the impact of the axioms on the simulated behavior of routing protocols, and provide detailed insight into the varied robustness of three popular radio models.

We recognize that analytical or simulation research in wireless networking must work with an abstraction of reality, modeling the behavior of the wireless network below the layer of interest. Unfortunately, overly simplistic or malformed assumptions can lead to misleading or incorrect conclusions.

Our results provide a counter-example to the notion that the arbitrary selection and generic configuration of any popular radio propagation model is sufficient for research on ad hoc routing algorithms. We do not claim to validate, or invalidate, the results of any other published study. Indeed, our point is that the burden is on the authors of past and future studies to a) clearly lay out the assumptions made in their simulation model, b) demonstrate whether those assumptions are reasonable within the context of their study, and c) clearly identify any limitations in the conclusions they draw.

4.1 The wireless network simulator

We used SWAN [LYN$^+$04], a simulator for wireless ad hoc networks that provides an integrated, configurable, and flexible environment for evaluating ad hoc routing protocols, especially for large-scale network scenarios. SWAN contains a detailed model of the IEEE 802.11 wireless LAN protocol and a stochastic radio channel model, both of which were used in this study.

We used SWAN's direct-execution simulation techniques to execute within the simulator the *same* routing code that was used in the experiments from the previous section. We modified the real routing code only slightly to allow multiple instances of a routing protocol implementation to run simultaneously in the simulator's single address space.

We extended the simulator to read the node mobility and application-level data logs generated by the real experiment. In this way, we were able to reproduce the same network scenario in simulation as in the real experiment. To further increase the fidelity of our simulation, we focused only on the 33 laptops that actually transmitted, received, and forwarded packets in the real experiments. To reproduce a comparable traffic pattern in simulation, the application traffic generator on each of the 33 active nodes still included the 7 crashed nodes as their potential packet destinations. Moreover, by directly running the routing protocols and the beacon service program, the simulator generated the same types of logs as in the real experiment. These conditions allow a direct comparison of results.

In the next few sections, we describe three simulation models with progressively unrealistic assumptions, and then present results to show the impact.

4.2 Our best model

We begin by comparing the results of the outdoor experiment with the simulation results obtained with our best signal propagation model and a detailed 802.11 protocol model. The best signal propagation model is a stochastic model that captures radio signal attenuation as a combination of two effects: small-scale fading and large-scale fading. Small-scale fading describes the rapid fluctuation in the envelope of a transmitted radio signal over a short period of time or a small distance, and primarily is caused by multipath effects. Although small-scale fading is in general hard to predict, wireless researchers over the years have proposed several successful statistical models for small-scale fading, such as the Rayleigh and Ricean distributions. Large-scale fading describes the slowly varying signal-power level over a long time interval or a large distance, and has two major contributing factors: distance path-loss and shadow fading. The distance path-loss models the average signal power loss as a function of distance: the receiving signal strength is proportional to the distance between the transmitter and the receiver raised to a given exponent. Both the free-space model and the two-ray ground reflection model mentioned earlier can be classified as distance path-loss models. The shadow fading describes the variations in the receiving signal power due to scattering; it can be modeled as a zero-mean log-normal distribution. Rappaport [Rap96] provides a detailed discussion of these and other models.

For our simulation, given the light traffic used in the real experiment, we used a simple SNR threshold approach instead of a more computationally intensive BER

[14] The two simpler models double the message delivery ratio predicted by the most detailed model. The even simpler analytical model used by STARA's designers would have exaggerated this value even more. One good simulation run, using a good stochastic model, would have revealed a drastic performance gap between their simple predictions and reality.

	Experiment	Simulation	Error
AODV	42.3%	46.8%	10.5%
APRL	17.5%	17.7%	1.1%
ODMRP	62.6%	56.9%	-9.2%

Table 2: Comparing message delivery ratios between real experiment and simulation.

approach. Under heavier traffic, this choice might have substantial impact [TMB01]. For the propagation model, we chose 2.8 as the distance path-loss exponent and 6 dB as the shadow fading log normal standard. These values, which must be different for different types of terrain, produce signal propagation distances consistent with our observations from the real network. Finally, for the 802.11 model, we chose parameters that match the settings of our real wireless cards.

Table 2 shows the difference in the overall message delivery ratio (MDR)—which is the total number of messages received by the application layer divided by the total number of messages generated—between the real experiment and the simulation. This propagation model produced relatively good results: the relative errors in predicted MDR were within 10% for all three routing protocols tested. We caution, however, that one cannot expect consistent results when generalizing this stochastic radio propagation model to deal with all network scenarios. After all, this model assumes some of the axioms we have identified, including flat earth, omni-directional radio propagation, and symmetry. In situations where such assumptions are clearly mistaken—for example, in an urban area—we should expect the model to deviate further from reality. Moreover, the real routing experiment provides a single reference point, and we do not have sufficient data to assess the overall effectiveness of the model under different network conditions.

On the other hand, since the model produced good results amenable to our particular outdoor experiment scenario, we use it in this study as the baseline to quantify the effect of the axioms on simulation studies. As we show, the axiom assumptions can significantly undermine the validity of the simulation results.

4.3 Simpler models

Next we weakened our simulator by introducing a simpler signal propagation model. We used the distance path-loss component from the previous model, but disabled the variations in the signal receiving power introduced by the stochastic processes. Note that these variations are a result of two distinct random distributions: one for small-scale fading and the other for shadow fading. The free-space model, the two-ray ground reflection model, and the generic distance path-loss model with a given exponent—all used commonly by wireless network researchers—differ primarily in the maximum distance that a signal can travel. For example, if we assume that the signal transmission power is 15 dBm and the receiving threshold is -81 dBm, the free-space model has a maximum range of 604 meters, the two-ray ground reflection model a range of 251 meters, and the generic path-loss model (with an exponent of 2.8) a range of only 97 meters. Indeed, we found that the receiving range plays an important role in ad hoc routing: longer distance shortens the data path and can drastically change the routing maintenance cost [LYN$^+$04].

In this study, we chose to use the two-ray ground reflection model since its signal travel distance matches observations from the real experiment.[15] This weaker model assumes Axiom 4: "If I can hear you at all, I can hear you perfectly," and specifically the testable axiom "The reception probability distribution over distance exhibits a sharp cliff." Without variations in the radio channel, all signals travel the same distance, and successful reception is subject only to the state of interference at the receiver. In other words, the signals can be received successfully with probability 1 as long as no collision occurs during reception.

Finally, we consider a third model that further weakens the simulator by assuming that the radio propagation channel is *perfect*. That is, if the distance between the sender and the recipient is below a certain threshold, the signal is received successfully with probability 1; otherwise the signal is always lost. The perfect-channel model represents an extreme case where the wireless network model introduces no packet loss from interference or collision, and the reception decision is based solely on distance. To simulate this effect, we bypassed the IEEE 802.11 protocol layer within each node and replaced it with a simple protocol layer that calculates signal reception based only on the transmission distance.

4.4 The results

First, we look at the reception ratio of the beacon messages, which were periodically sent via broadcasts by the beacon service program on each node. We calculate the reception ratio by inspecting the entries in the beacon logs, just as we did for the real experiment. Figure 22 plots the beacon reception ratios during the execution of the AODV routing protocol. The choice of routing protocol is unimportant in this study since we are comparing

[15] When we consider the full experiment field, which provides possible reception ranges of over 500 meters, we see almost no receptions beyond 250 meters. The 251-meter range of the 2-ray model is computed from a well-known formula, using a fixed transmit power (15 dBm) and antenna height (1.0 meter).

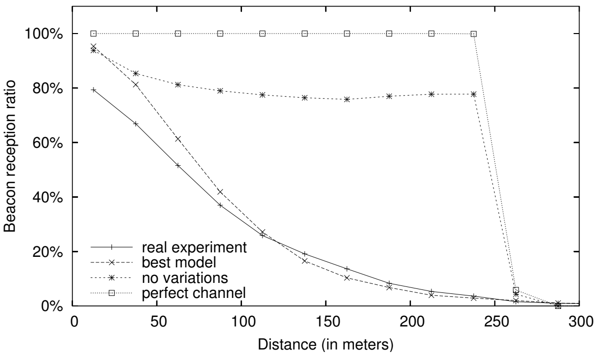


Figure 22: The beacon reception ratio at different distances between the sender and the receiver. The probability for each distance bucket is plotted as a point at the midpoint of its bucket; this format is easier to read than the boxes used in earlier plots.

the results between the real experiment and simulations. We understand that the control messages used by the routing protocol may slightly skew the beacon reception ratio due to the competition at the wireless channel.

Compared with the two simple models, our best model is a better fit for the real experiment results. It does, however, slightly inflate the reception ratios at shorter distances and underestimate them at longer distances. More important for this study is the dramatic difference we saw when signal power variations were not included in the propagation model. The figure shows a sharp cliff in the beacon reception ratio curve: the quality of the radio channel changed abruptly from relatively good reception to zero reception as soon as the distance threshold was crossed. The phenomenon is more prominent for the perfect channel model. Since the model had no interference and collision effects, the reception ratio was 100% within the propagation range.

Next, we examine the effect of different simulation models on the overall performance of the routing protocols. Figures 23–25 show the message delivery ratios, for the three ad hoc routing algorithms, as we varied the application traffic intensity by adjusting the average message inter-arrival time at each node. Note the logarithmic scale for the x-axes in the plots. The real experiment's result is represented by a single point in each plot.

Figures 23–25 show that the performance of routing algorithms predicted by different simulation models varied dramatically. For AODV and APRL, both simple models exaggerated the message delivery ratio *significantly*. In those models, the simulated wireless channel was much more resilient to errors than the real network, since there were no spatial or temporal fluctuations in signal power. Without variations, the transmissions had a much higher chance to be successfully received, and in turn, there were fewer route invalidations, and more packets were able to find routers to their intended destinations. The performance of the perfect-channel model remained insensitive to the traffic load since the model did not include collision and interference calculations at the receiver, explaining the divergence of the two simple models as the traffic load increases. For ODMRP, we cannot make a clear distinction between the performance of the best model and of the no-variation model. One possible cause is that ODMRP is a multicast algorithm and has a more stringent bandwidth demand than the strictly unicast protocols. A route invalidation in ODMRP triggers an aggressive route rediscovery process, and could cause significant packet loss under any of the models.

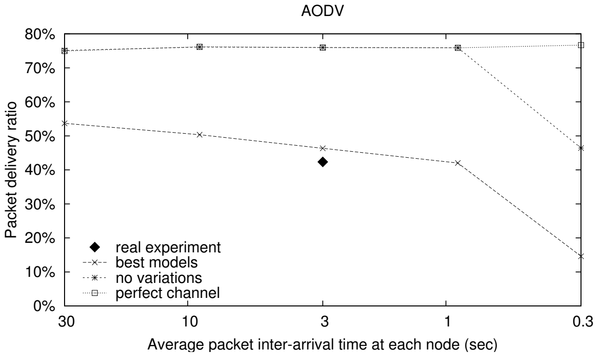


Figure 23: Message delivery ratios for AODV.

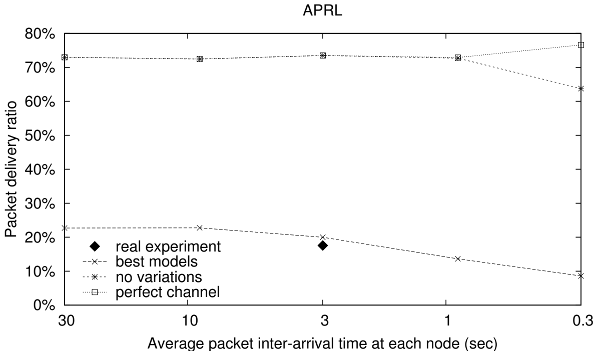


Figure 24: Message delivery ratios for APRL.

In summary, the assumptions embedded inside the wireless network model have a great effect on the simulation results. On the one hand, our best wireless network model assumes some of the axioms, yet the results do not differ significantly from the real experiment results. On the other hand, one must be extremely careful when assuming some of the axioms. If we had held our experiment in an environment with more hills or obstacles, the simulation results would not have matched as well. Even in this relatively flat environment, our study shows

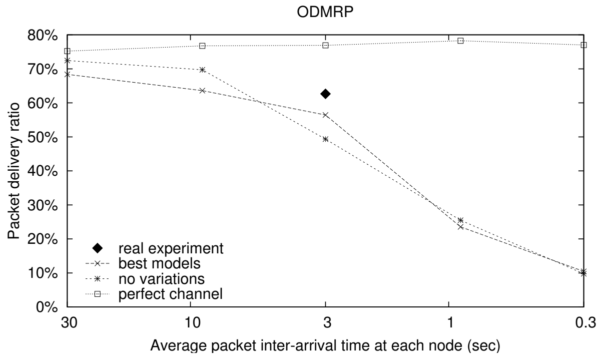


Figure 25: Message delivery ratios for ODMRP.

that proper modeling of the lossy characteristics of the radio channel has a significant impact on the routing protocol behaviors. For example, using our best model, one can conclude from Figure 23 and Figure 25 that ODMRP performed better than AODV with light traffic load (consistent with real experiment), but that their performance was comparable when the traffic was heavy. If we use the model without variations, however, one might arrive at the opposite conclusion, that AODV performed consistently no worse than ODMRP. The ODMRP results are interesting by themselves, since the packet-delivery degradation as the traffic load increases is more than might be expected for an algorithm designed to find redundant paths (through the formation of appropriate forwarding groups). Bae has shown, however, that significant degradation can occur as intermediate nodes move, paths to targets are lost, and route rediscovery competes with other traffic [BLG00]. In addition, the node density was high enough that each forwarding group could have included a significant fraction of the nodes, leading to many transmitted copies of each data packet. An exploration of this issue is left for future work.

4.5 Further investigation

In the previous sections, we investigate the impact of assuming the axioms. We demonstrate that certain assumptions dramatically affect the results, and in some cases even reverse the ranking of the algorithms being compared. Accordingly, we conclude that simulation designers should be wary of what assumptions they make in constructing their models. In this section we extend our investigation with a series of related simulation experiments.

Specifically, we combined three common radio propagation models with the connectivity trace derived from the outdoor experiment beacon logs, leading to six different radio propagation models in simulation: three using the connectivity traces and the other three not. In the first three cases, we used the connectivity trace to determine whether a packet from a mobile station could reach another mobile station, and then we used the radio propagation models to determine the receiving power for the interference calculation. Comparison of models with measured connectivity with those without give us a means of refining a model's power—if a model is seen to require connectivity information to work well, it is not a robust model because the power of prediction comes from measurements. On the other hand if a model without measured connectivity information works about as well as does the version with it, then the model itself contains accurate predictive power for connectivity.

Recognizing that message delivery ratio is not the only metric of interest to routing protocol designers, we also describe the performance of our six models in generating accurate hop count distributions.

We then investigate the sensitivity of two important large-scale fading parameters: the distance path-loss exponent, and the standard deviation value for shadow fading. The choice of these parameters is often arbitrary in simulation studies, and we maintain that these values are important determinents of the environment being simulated, and therefore should be selected and accounted for with care.

The three models used in this further investigation—generic propagation, two-ray ground reflection, and Friis free-space—are similar, but not exactly the same as the models used in quantifying the impact of the axioms.

Our generic propagation model is the same as the best model from the axioms investigation. It uses the same large-scale and small-scale fading models, and the same parameter values for the distance path-loss exponent and shadow-fading standard deviation. Similarly, the two-ray ground reflection model is the same as the first weakening of the axiom's best model. It uses no shadow-fading or small-scale fading, and is configured with the same path-loss exponent as its axioms investigation equivalent.

The Friis free-space model, however, does differ from the perfect channel model used previously. The perfect model had no implementation of an 802.11 protocol, and instead used only a simple distance threshold to model packet reception. The Friis free-space model, on the other hand, retains an implementation of the 802.11 protocol, allowing for collisions and interference. And though it assumes an ideal radio propagation condition—the signals travel in a vacuum space without obstacles—the power loss is proportional to the square of the distance between the transmitter and the receiver. Because the Friis model uses signal strength to calculate collisions, this use of a path-loss exponent is a more complicated estimation of radio behavior than the simple distance threshold used in the perfect model.

We choose this slightly more complicated model for this further investigation because the previous sections

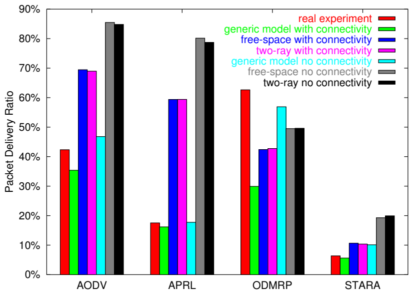


Figure 26: Comparing the message delivery ratio from the real experiment with various radio propagation models. "With connectivity" means the connectivity trace was used.

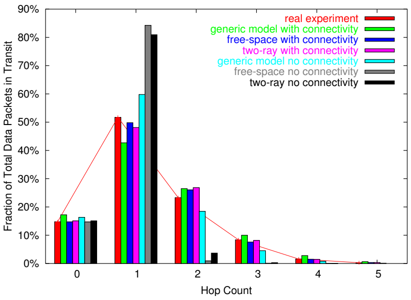


Figure 27: The hop-count histogram of AODV in real experiment and in simulation.

make it clear that the perfect model, without any compensation for collisions or interference, grossly overestimates protocol performance in all cases. The use of the Friis free-space model in this experiment provides more interesting results by allowing us to compare three more related approaches to estimating radio reception power over time and distance.

It should also be noted that to further increase the similarity between the simulated environment and the real conditions, we modified the application traffic generator to read the outdoor experiment application log and generate the same packets as in the real experiment. We were unable to implement this feature in the axiom experiments because we ran those simulations at various rates of packet generation.

Results. We first examine the message delivery ratio. Figure 26 shows the message delivery ratio from the real experiment and the simulation runs with six radio propagation models (three of which used the connectivity trace derived from the real experiment to determine the reachability of the signals). Each simulation result is an average of five runs; the variance is insignificant and therefore not shown.

These results verify many of the conclusions reached in the previous simulation experiment. For example, the generic propagation model, with typical parameters, offers an acceptable prediction of the routing algorithm performance. Different propagation models predict vastly different protocol behaviors, and these differences are non-uniform across the algorithms tested. For three algorithms, the two-ray ground reflection and free-space models both exaggerate the PDR, whereas ODMRP performance was underestimated.

More important is what we observe by comparing the models with connectivity traces to those without. The propagation models that used the connectivity trace in general lower the message delivery ratio, when comparing with the propagation models that do not use the connectivity trace. This result is not surprising: the connectivity trace, to some degree, can represent the peculiar radio propagation scenario of the test environment. In our experiment, there were significant elevation changes in the test field that led to the obstruction of radio signals between laptops that were close by in distance. Without connectivity traces, the propagation model assumes an omni-directional path loss dependent only on the distance, which resulted in a more connected network (fewer hops) and therefore better delivery ratio.

Of course, the message delivery ratio does not reflect the entire execution environment of the routing algorithm. From the routing event logs, we collected statistics related to each particular routing strategy. Figure 27 shows a histogram of the number of hops that a data packet traversed in AODV, before it either reached its destination or dropped along the path. For example, a hop count of zero means that the packet was dropped at the source node; a hop count of one means the packet went one hop: either the destination was its neighbor or the packet failed to reach the next hop. The figure shows the fraction of the data packets that traveled in the given number of hops. As above, the free-space and two-ray models resulted in fewer hops by exaggerating the transmission range. We also see that the connectivity trace was helpful in predicting the hop counts, which confirms that the problem with the free-space and two-ray models using the connectivity trace was that they did not consider packet losses due to the variations in receiving signal power.

Finally, we take a look at the sensitivity of certain simulation parameters in the generic propagation model. The exponent for the distance path loss and the standard deviation in log-normal distribution for the shadow fading are heavily dependent on the environment under investigation. In the next experiment, we ran a simulation with the same number of mobile stations and with the same traffic load as in the real experiment. Figure 28 shows AODV performance in packet delivery ratio, as we varied the path-loss exponent from 2 to 4 and the shadow log-

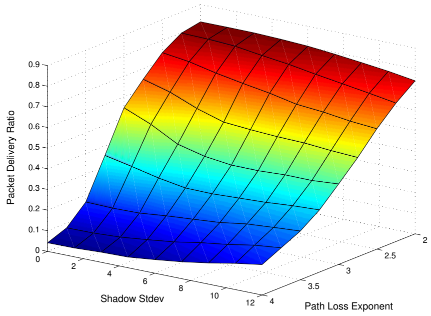


Figure 28: Sensitivity of the AODV protocol performance to the parameters of the large-scale fading model.

normal standard deviation from 0 to 12 dB—the ranges suggested by [Rap96] for radio modeling out-of-doors.

The AODV behavior was more sensitive to the path-loss exponent than to the shadow standard deviation. That is, the signal propagation distance had a stronger effect on the algorithm's performance. A shorter transmission range means packets must travel through more hops (via longer routes) before reaching its destination, and therefore has a higher probability to be dropped. A larger shadow standard deviation caused the links to be more unstable, but the effect varied. On the one hand, when the path-loss exponent was small, the signals had a long transmission range, so the small variation in the receiving signal strength did not have a significant effect on routing, causing only infrequent link breakage. On the other hand, when the exponent was large, most nodes were disconnected. A variation in the receiving signal power helped establish some routes that were impossible if not for the signal power fluctuation. Between the extremes, a larger variation in the link quality generally caused more transmission failures, and therefore resulted in slightly lower message delivery ratio.

The critical implication of this sensitivity study is that we cannot just grab a set of large-scale fading parameters, use them, and expect meaningful results for any specific environment of interest. On the one hand, pre-simulation empirical work to estimate path-loss characteristics might be called for, if the point of the experiment is to quantify behavior in a given environment. Alternatively, one may require more complex radio models (such as ray tracing) that include complex explicit representations of the domain of interest. On the other hand, if the objective is to compare protocols, knowledge that the generic propagation model is good lets us compare protocols using a range of path-loss values. While this does not *quantify* behavior, it may allow us to make *qualitative* conclusions about the protocols over a range of environments.

To summarize, this further investigation reaffirms our earlier conclusion that it is critical to choose a proper wireless model that reflects a real-world scenario for studying the performance of ad hoc routing algorithms. In contrast to earlier studies [TBTG01], we found that using a simple stochastic radio propagation model with parameters typical to the outdoor environment can produce acceptable results. We must recognize, however, the results are sensitive to these parameters. It is for this reason we caution that the conclusions drawn from simulation studies using simple propagation models should apply only to the environment they represent. The free-space model and the two-ray model, which exaggerate the radio transmission range and ignore the variations in the receiving signal power, can largely misrepresent the network conditions.

5 Conclusion

In recent years, dozens of Mobicom and Mobihoc papers have presented simulation results for mobile ad hoc networks. The great majority of these papers rely on overly simplistic assumptions of how radios work. Both widely used radio models, "flat earth" and `ns-2` "802.11" models, embody the following set of axioms: the world is two dimensional; a radio's transmission area is roughly circular; all radios have equal range; if I can hear you, you can hear me; if I can hear you at all, I can hear you perfectly; and signal strength is a simple function of distance.

Others have noted that real radios and ad hoc networks are much more complex than the simple models used by most researchers [PJL02], and that these complexities have a significant impact on the behavior of MANET protocols and algorithms [GKW$^+$02]. In this study, we validate the importance of this problem, and present results and recommendations to help researchers generate more reliable simulations.

We present the results of an unprecedented large-scale outdoor experiment comparing in detail the performance of four different ad hoc algorithms. We then enumerate the set of common assumptions used in MANET research, and use data from our real-world experiment to strongly contradict these "axioms." Finally, we describe a series of simulation experiments that quantify the impact of assuming these axioms and validate a group of radio models commonly used in ad hoc network research.

The results cast doubt on published simulation results that implicitly rely on our identified assumptions, and provide guidance for designing and configuring more reliable simulation models for use in future studies.

We conclude with a series of recommendations, ...for the MANET research community:

1. Choose your target environment carefully, clearly list your assumptions about that environment, choose simulation models and conditions that match those as-

sumptions, and report the results of the simulation in the context of those assumptions and conditions.

2. Use a realistic stochastic model when verifying a protocol, or comparing a protocol to existing protocols. Furthermore, any simulation should explore a range of model parameters since the effect of these parameters is not uniform across different protocols. Simple models are still useful for the initial exploration of a broad range of design options, due to their efficiency.
3. Consider three-dimensional terrain, with moderate hills and valleys, and corresponding radio propagation effects. It would be helpful if the community agreed on a few standard terrains for comparison purposes.
4. Include some fraction of asymmetric links (e.g., where A can hear B but not vice versa) and some time-varying fluctuations in whether A's packets can be received by B or not. Here the `ns-2` "shadowing" model may prove a good starting point.
5. Use real data as input to simulators, where possible. For example, using our data as a static "snapshot" of a realistic ad hoc wireless network with significant link asymmetries, packet loss, elevated nodes with high fan-in, and so forth, researchers could verify whether their protocols form networks as expected, even in the absence of mobility. The dataset also may be helpful in the development of new, more realistic radio models.
6. Recognize that connectivity is easily overestimated in simulation. Even the most realistic models used in our experiments overestimated network connectivity as compared to our real-world results. This observation is especially important when validating or comparing a protocol that depends on a certain minimum threshold of network connections to perform effectively.
7. Avoid simple models, such as free-space or two-ray ground reflection, when validating or comparing a protocol for which hop count is a vital component of its performance. These models significantly exaggerate transmission range, and subsequently lower hop counts to an unrealistic level.

...for simulation and model designers:

1. Allow protocol designers to run the same code in the simulator as they do in a real system [LYN$^+$04], making it easier to compare experimental and simulation results.
2. Develop a simulation infrastructure that encourages the exploration of a range of model parameters.
3. Develop a range of propagation models that suit different environments, and clearly define the assumptions underlying each model. Models encompassing both physical and data-link layer need to be especially careful.
4. Support the development of standard terrain and mobility models, and formats for importing real terrain data or mobility traces into the simulation.

...for protocol designers:

1. Consider carefully your assumptions of lower layers. In our experimental results, we found that the success of a transmission between radios depends on many factors (ground cover, antenna angles, human and physical obstructions, background noise, and competition from other nodes), most of which cannot be accurately modeled, predicted or detected at the speed necessary to make per-packet routing decisions. A routing protocol that relies on an acknowledgement quickly making it from the target to the source over the reverse path, that assumes that beacons or other broadcast traffic can be reliably received by most or all transmission-range neighbors, or that uses an instantaneous measure of link quality to make significant future decisions, is likely to function significantly differently outdoors than under simulation or indoor tests.
2. Develop protocols that adapt to environmental conditions. In our simulation results, we found that the relative performance of two algorithms (such as AODV and ODMRP) can change significantly, and even reverse, as simulation assumptions or model parameters change. Although some assumptions may not significantly affect the agreement between the experimental and simulation results, others may introduce radical disagreement. For similar reasons, a routing protocol tested indoors may work very differently outdoors. Designers should consider developing protocols that make few assumptions about their environment, or are able to adapt automatically to different environmental conditions.
3. Explore the costs and benefits of control traffic. Both our experimental and simulation results hint that there is a tension between the control traffic needed to identify and use redundant paths and the interference that this extra traffic introduces when the ad hoc routing algorithm is trying to react to a change in node topology. The importance of reducing interference versus identifying redundant paths (or reacting quickly to a path loss) might appear significantly different in real experiments than under simple simulations, and protocol designers must consider carefully whether extra control traffic is worth the interference price.
4. Use detailed simulation as a tool to aid the protocol design process. Modeling the effects of collisions and highly variant transmission strengths may provide some guidance for tailoring your protocol design to more effectively avoid or adapt to destabilizing environmental conditions.

Availability. We will make our simulator and our dataset available to the research community upon completion of a conference version of this paper. The dataset, including the actual position and connectivity measurements, would be valuable as input to future simulation experiments. The simulator contains several radio-propagation models.

Acknowledgments. This thesis contains some text, data, and analysis adapted from two of our related studies. Accordingly, I would like to explicitly thank and acknowledge all of the relevant co-authors. Section 2.2, Section 3, Section 4, and Section 5 use material derived from *The mistaken axioms of wireless-network research*, which is an update of an existing technical report [KNE03], and is currently under conference submission. The co-authors for this work are: David Kotz, Robert Gray, Chip Elliott, Calvin Newport, Jason Liu, and Yougu Yuan. Section 2.1 and Section 4.5 use material derived from *Simulation Validation Using Direct Execution of Wireless Ad-Hoc Routing Protocols* [LYN$^+$04], which was be published in the proceedings of the PADS '04 conference. The co-authors for this work are: Jason Liu, Yougu Yuan, David Nicol, Robert Gray, Calvin Newport, David Kotz, and Luiz Felipe Perrone.

I would also like to thank my thesis committee of David Kotz, Bob Gray, and Chris McDonald for their insightful comments. In particular, I thank Bob Gray for his help in bringing together the outdoor experiment section. And I want to especially thank my advisor David Kotz for all of his support and guidance over the last two years.

This work was supported in part by the Dartmouth Center for Mobile Computing, DARPA (contract N66001-96-C-8530), the Department of Justice (contract 2000-CX-K001), the Department of Defense (MURI AFOSR contract F49620-97-1-03821), and the Office for Domestic Preparedness, U.S. Department of Homeland Security (Award No. 2000-DT-CX-K001). Additional funding provided by the Dartmouth College Dean of Faculty office in the form of a Presidential Scholar Undergraduate Research Grant, and a Richter Senior Honors Thesis Research Grant. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security, the Department of Justice, the Department of Defense, or any other branch of the U.S. Government.

References

[BLG00] Sang Ho Bae, Sung-Ju Lee, and Mario Gerla. Unicast performance analysis of the ODMRP in a mobile ad-hoc network testbed. In *ICCCN 2000*, October 2000.

[ER00] Moncef Elaoud and Parameswaran Ramanathan. Adaptive allocation of CDMA resources for network-level QoS assurances. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, pages 191–199. ACM Press, 2000.

[FV02] Kevin Fall and Kannan Varadhan. *The ns Manual*, April 14 2002. www.isi.edu/nsnam/ns/ns-documentation.html.

[GK97] P. Gupta and P. R. Kumar. A system and traffic dependent adaptive routing algorithm for ad hoc networks. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 2375–2380, December 1997.

[GKW$^+$02] Deepak Ganesan, Bhaskar Krishnamachari, Alec Woo, David Culler, Deborah Estrin, and Stephen Wicker. Complex behavior at scale: An experimental study of low-power wireless sensor networks. Technical Report UCLA/CSD-TR 02-0013, UCLA Computer Science, 2002.

[Gra00] Robert S. Gray. Soldiers, agents and wireless networks: A report on a military application. In *Proceedings of the Fifth International Conference and Exhibition on the Practical Application of Intelligent Agents and Multi-Agents (PAAM 2000)*, Manchester, England, April 2000.

[Gud91] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 27(23):2145–2146, November 1991.

[Gup00] Piyush Gupta. *Design and Performance Analysis of Wireless Networks*. PhD thesis, Department of Electrical Engineering, University of Illinois at Urbana-Champaign, 2000.

[JLH$^+$99] Per Johansson, Tony Larsson, Nicklas Hedman, Bartosz Mielczarek, and Mikael Degermark. Scenario-based performance analysis of routing protocols for mobile ad-hoc networks. In *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networking*, pages 195–205. ACM Press, 1999.

[KK98] B. Karp and H. T. Kung. Dynamic neighbor discovery and loopfree, multi-hop routing for wireless, mobile networks. Harvard University, May 1998.

[KLM+00] O. E. Kelly, J. Lai, N. B. Mandayam, A. T. Ogielski, J. Panchal, and R. D. Yates. Scalable parallel simulations of wireless networks with WiPPET: Modeling of radio propagation, mobility and protocols. *Mobile Networks and Applications*, 5(3):199–208, September 2000.

[KNE03] David Kotz, Calvin Newport, and Chip Elliott. The mistaken axioms of wireless-network research. Technical Report TR2003-467, Dartmouth College, Computer Science, Hanover, NH, July 2003.

[Lee82] W. .C. Y. Lee. *Mobile communications engineering*. McGraw-Hill, 1982.

[LGC02] S. J. Lee, M. Gerla, and C. C Chiang. On-demand multicast routing protocol in multihop wireless mobile networks. *ACM/Kluwer Mobile Networks and Applications, special issue on Multipoint Communications in Wireless Mobile Networks*, 7(6):441–453, December 2002.

[Lun02] David Lundberg. Ad hoc protocol evaluation and experiences of real world ad hoc networking. Master's thesis, Department of Information Technology, Uppsala University, Sweden, 2002.

[LYN+04] Jason Liu, Yougu Yuan, David M. Nicol, Robert S. Gray, Calvin C. Newport, David F. Kotz, and Luiz Felipe Perrone. Simulation validation using direct execution of wireless ad-hoc routing protocols. In *18th Workshop on Parallel and Distributed Simulation (PADS'04)*, May 2004.

[PJL02] K. Pawlikowski, H.-D.J Jeong, and J.-S.R. Lee. On credibility of simulation studies of telecommunication networks. *IEEE Communications*, 40(1):132–139, January 2002.

[PR99] C. E. Perkins and E. M. Royer. Ad hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pages 90–100, February 1999.

[Rap96] T. S. Rappaport. *Wireless Communications, Principles and Practice*. Prentice Hall, Upper Saddle River, New Jersey, 1996.

[TBTG01] M. Takai, R. Bagrodia, K. Tang, and M. Gerla. Efficient wireless network simulations with detailed propagation models. *Wireless Networks*, 7(3):297–305, May 2001.

[TMB01] Mineo Takai, Jay Martin, and Rajive Bagrodia. Effects of wireless physical layer modeling in mobile ad hoc networks. In *Proceedings of MobiHoc 2001*, pages 87–94. ACM Press, 2001.

[YLA02] W. H. Yuen, H. Lee, and T. Andersen. A simple and effective cross layer networking system for mobile ad hoc networks. *13th IEEE Int'l Symposium on Personal, Indoor and Mobile Radio Communications*, September 2002.

[ZL02] Yongguang Zhang and Wei Li. An integrated environment for testing mobile ad-hoc networks. In *Proceedings of MobiHoc 2002*, pages 104–111. ACM Press, 2002.