Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Ad Hoc Networks 3 (2005) 762–776

www.elsevier.com/locate/adhoc

Topology and mobility considerations in mobile ad hoc networks

Brent Ishibashi, Raouf Boutaba *

School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1

Received 1 September 2003; accepted 1 March 2004
Available online 28 April 2004

Abstract

A highly dynamic topology is a distinguishing feature and challenge of a mobile ad hoc network. Links between nodes are created and broken, as the nodes move within the network. This node mobility affects not only the source and/or destination, as in a conventional wireless network, but also intermediate nodes, due to the network's multihop nature. The resulting routes can be extremely volatile, making successful ad hoc routing dependent on efficiently reacting to these topology changes.

In order to better understand this environment, a number of characteristics have been studied concerning the links and routes that make up an ad hoc network. Several network parameters are examined, including number of nodes, network dimensions, and radio transmission range, as well as mobility parameters for maximum speed and wait times. In addition to suggesting guidelines for the evaluation of ad hoc networks, the results reveal several properties that should be considered in the design and optimization of MANET protocols.

Keywords: Ad hoc networks; Topology; Characteristics; Mobility; Stability; Path lifetime

1. Introduction

The mobile devices in a mobile ad hoc network (MANET) play a very different role than in a conventional wireless LAN (WLAN). In a conventional WLAN, communications are centered on the base station or access point; the infrastructure up to the base station is mostly fixed, so the topology is stable. In a MANET, mobile nodes act not only as end systems, but also as routing devices. The topology of the network is dependent on the relative locations and connections of nodes within the network.

This results in a topology that is potentially extremely dynamic. This effects all aspects of an

---

* Corresponding author. Tel.: +1 519 888 4567x5897; fax: +1 519 885 1208.

E-mail addresses: bkishiba@uwaterloo.ca (B. Ishibashi), rboutaba@uwaterloo.ca (R. Boutaba).

1570-8705/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2004.03.013

ad hoc network, including the medium access control (MAC) layer and routing protocols. In order to achieve acceptable performance, the MANET as a whole must find effective ways of managing the side caused by the changing topology. In order to achieve this, those effects must first be understood.

To date, most research has focused on practical aspects—development of mechanisms and protocols for use in ad hoc networking, particularly for routing. This has reached limited success. Although the concept of a MANET has been shown to be workable, in practice the performance has been inadequate to gain widespread acceptance or commercial feasibility. In fact, even the understanding of the resulting protocols is questionable, due to the difficulty of comparing evaluation results.

A better knowledge of the effects of various MANET parameters and characteristics will greatly aid in the development of new ideas. A better understanding will:

- Allow a wiser selection of scenarios for testing and evaluation.
- Reduce or eliminate design decisions based on faulty assumptions.
- Promote optimization of parameters for real network conditions.
- Reveal characteristics to be used in developing future protocols.

With these potential benefits in mind, a number of experiments have been performed in order to investigate the effects of several parameters, including the number of nodes, their behaviour, and the environment in which they exist. Several aspects have been explored, including the overall connection characteristics in the network, the severity of the topology instability, as well as the resulting effects on routing in a MANET.

2. Background

2.1. Mobile ad hoc networks

A mobile ad hoc network consists of a collection of wireless-enabled devices. Links are formed between node-pairs that are within direct communication range, and the nodes and links combine to create the network topology. During the lifetime of the network, nodes may move around within the network, altering the topology by creating or breaking links between nodes.

Nodes may also enter or leave the network. This may be due to mobility, if a node moves out of range of all other nodes in the network. This occurs most frequently near the geographical edge of the network cluster of nodes as in Fig. 1. Alternatively, a node may enter or leave the network. Switching a device off voluntarily removes the node from the network (and the routing process), however a similar effect could be caused by a node failure, such as a node's loss of power.

As direct transmission range is limited by radio propagation effects, particularly attenuation, a MANET cannot form links with all other nodes in the network. Instead, all a small set of nodes are reachable, the neighbour set. Traffic from a source, destined for a node not in its neighbour set, must be forwarded in a multihop manner. The source sends the packet to one of its neighbours, who in turn forwards it to another neighbour, until the packet reaches the destination node, as seen in Fig. 2. In order for this to occur, the path from source to destination must be determined.

2.2. Types of MANET routing

In the most basic form, nodes could operate as repeaters; this was the basis for initial efforts in

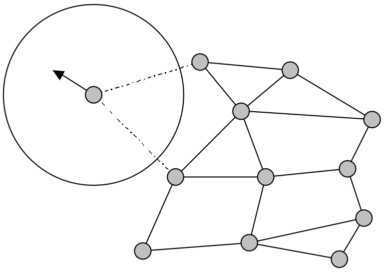

Fig. 1. Disconnection of a node near the MANET edge. The node has moved out of range, and can no longer reach the rest of the network.

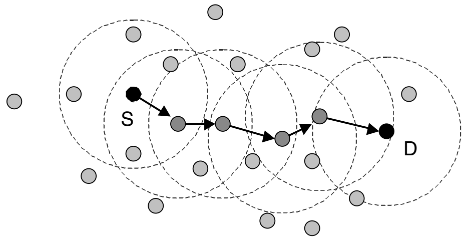

Fig. 2. Multihop traffic delivery of a MANET. Packets are transmitted from node to node until they reach their destination.

packet radio systems, the forerunner of ad hoc networking. However, this is not an efficient solution for all MANET traffic. Instead, a sequence of forwarding nodes must be determined. The packet is then relayed directly from one intermediate node to the next.

Three classifications of MANET routing protocols have emerged. All three types have the same ultimate goal: to have a correct route available when a source wishes to send to a destination, however each achieve this in a different manner. Due to the differences in their approaches, each has its own strengths and weaknesses, and variations in network characteristics affect them in different manners. Therefore, it is a good idea to be aware of the current approaches in discussing MANET characteristics.

2.2.1. Proactive routing

Proactive routing algorithms, also described as table-driven algorithms, closely resemble many traditional wired routing algorithms. Each node maintains a routing table, containing routing information on reaching every other node in the network. The routing information is always available. If a node wishes to send, it looks up the routing information and sends the packet.

In order to have this routing information available and up to date, routing information must be exchanged. In a wired network, where topology changes are relatively infrequent, these updates are exchanged whenever a change is detected. Unfortunately, topology changes in a MANET occur far more often. Therefore, a variety of strategies were used to delay or minimize the sending of updates, in order to reduce the overhead incurred by the updates.

Several of the earliest ad hoc routing protocols used the proactive approach. Destination-sequenced distance-vector routing (DSDV) [1] adapted the classical Bellman–Ford algorithm [2] with methods for reducing the update frequency. Clusterhead gateway switch routing (CGSR) [3] added clustering in order to reduce overhead and improve scalability. More recently, optimized link state routing (OLSR) [4] was introduced, exchanging link state information rather than routes.

2.2.2. Source-initiated on-demand routing

On-demand routing takes a very different approach. Rather than building routes in advance, these protocols take a reactive approach. When a source wishes to send to a particular destination, it initiates the route discovery process, in order to find the destination. Upon reaching the destination, the destination replies to the source, creating the route.

Unlike in a proactive protocol, routes are not immediately available for a node to use. A delay is incurred whenever a route has to be found, both at the initial discovery, as well as any time the route fails and has to be rediscovered. In addition, this discovery usually requires some type of broadcast in order to find the destination. However, only those routes that are needed are created, reducing the total amount of protocol overhead.

Ad hoc on-demand distance vector (AODV) [5] and dynamic source routing (DSR) [6] are the two primary examples of on-demand protocols. In AODV, next hop information is stored at each intermediate node along the path, set up as the reply packet is passed backwards along the selected path. DSR returns the entire path to the source, and complete path information is included by the source in all subsequent data packets.

2.2.3. Hybrid protocols

A hybrid class of protocols has also emerged, combining the features of proactive and on-demand types. These protocols attempt to minimize the weaknesses of the respective classes. Zone routing protocol (ZRP) [7] uses a two part approach: a node uses a proactive approach to keep routes for its local zone (nodes within a certain number of hops), while more distant destinations are found using a reactive discovery process. Therefore, a hybrid protocol's characteristics would also be a combination of the properties of proactive and on-demand protocols.

2.2.4. Shortest paths

Routing must not only find a route, but in many cases, it must choose between a number of different routes. With the exception of a few power-based [8–10] or link/signal stability [11,12] routing algorithms, most MANET routing protocols focus on minimizing the hop count of the chosen route. There are a number of reasons why this is important in a multihop wireless network.

The number of hops corresponds with the number of times a packet must be successfully transmitted and received in order to reach its destination. In a wireless network, each additional transmission has a number of consequences. First, wireless communication uses a shared medium, so each transmission makes the medium busy for the length of the transmission. A longer path consumes additional bandwidth, preventing other nodes from transmitting. Second, additional hops add to the delay experienced by a packet, due to the additional buffering, contention, and transmission time required.

There is sometimes an argument for choosing a longer path. The shortest hop path may tend to be susceptible to an "edge effect" [13]. When each node along the path is near the maximum transmission range of the preceding node, any node mobility is likely to break a link (Fig. 3). Choosing a path with more hops (but still in a straight line towards the destination) may improve the lifetime of the path (Fig. 4). However, simply choosing a longer path is not sufficient, as a non-straight line path may still experience edge transmission effects.

In terms of overall network throughput, the shorter path should be better. However, choosing a path with more stable links may reduce the demands on the routing protocol, resulting in improved real performance. In addition, other factors, such as energy or congestion levels of nodes along the path, may also affect the choice of the best path.

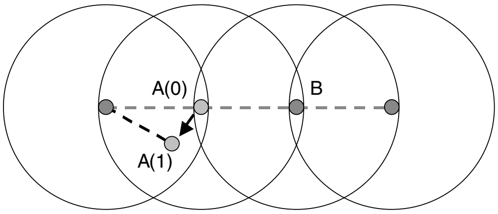

Fig. 3. The "edge effect". Each hop covers a near maximum transmission length in a straight line towards the destination. A small movement by any node may break the path. Here, the movement of A at $t = 1$ severs the link between A and B.

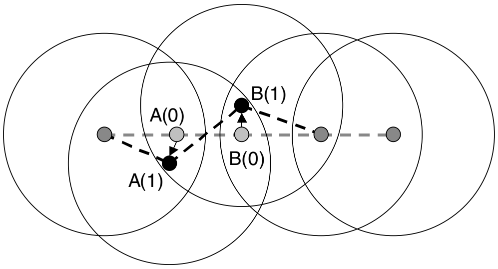

Fig. 4. A longer, more stable path. By choosing a longer path, with each hop well below the maximum transmission length, nodes have a greater freedom of movement without breaking the path. Nodes A and B can now move (to a limited degree) without breaking the path.

2.2.5. Multiple paths

One consequence of every node in the network participating in the routing process is the presence of a large number of similar paths. In a network with a sufficient density of nodes (each node connecting with a number of others), any particular source–destination pair may have a number of paths. This may include some of paths of length equal to the shortest path, as well as a number of longer paths (Fig. 5).

In many routing protocols, the particular path to use is selected based on the order of arrival of routing packets. However, a number of ideas have been proposed to make better use of these multiple paths. Using multiple routes can improve the probability of maintaining a working route, or better spread the traffic load over a larger number of nodes [14]. The availability and use of multiple paths could help a network better adjust to the loss or movement of a particular node.

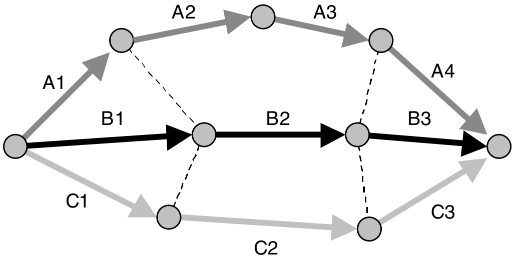

Fig. 5. Multiple paths (A, B, and C) between a source and destination.

2.3. *Delivering packets*

While routing takes care of finding and maintaining the required routes, delivery of packets still requires the packets to be transmitted by the source and retransmitted at each intermediate node along the path. Compared to a conventional wireless network, where a single transmission is normally used to deliver a packet from the mobile to the base station, an ad hoc packet may utilize the wireless medium multiple times before the packet reaches the destination. Each transmission consumes valuable bandwidth and power, two resources that are already scarce in a wireless network.

Nodes must contend for the medium with other nodes within transmission range, that is their neighbours. Therefore, they must share the available bandwidth with those neighbours. If a node has a large number of neighbours who also require the medium, the node will have less opportunity to send packets, and it will be forced to wait longer in order to send its packets. Heavy contention for the medium tends to stress the MAC protocol, leading to many collisions and inefficient performance.

3. Network scenarios

The network scenarios to be examined were chosen to mimic the common scenarios chosen to test MANET routing protocols. Three different variables were provided to examine static network scenarios, with two additional variables added for mobile network scenarios.

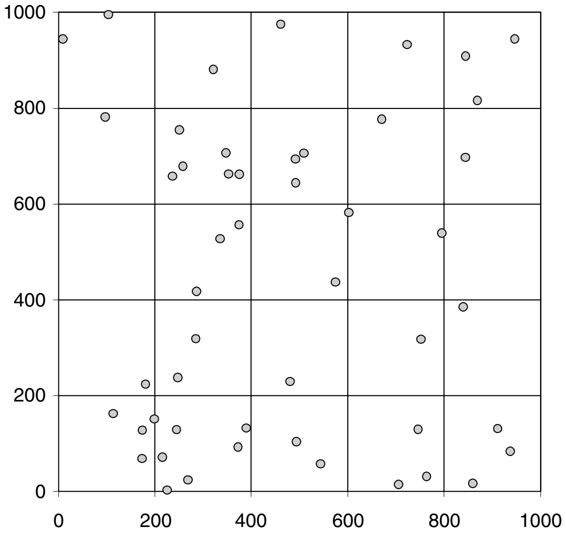

Fig. 6. A MANET configuration. A network instance with parameters $N = 50$, $S = 1000$ m.

3.1. *Static conditions*

In each case, the network existed within a square area with sides of length S. The nodes of the network moved inside this area, remaining within and active at all times. N nodes were placed randomly at points within this area, using a uniform distribution. All nodes were considered to be co-planar, and existing in free space. An example configuration is depicted in Fig. 6 with $S = 1000$ and $N = 50$.

A radio transmission range R was chosen. For the purpose of these experiments, this was a nominal range, with no variation. A link could be established successfully between any two nodes that were located within a distance of R from each other. Beyond this distance, no direct link could exist. No link layer effects, such as HELLO packet losses were considered, although this has been shown to have a real effect on the establishment of links in a MANET [15]. The links can be represented by edges connecting adjacent nodes, creating a graph topology.

3.2. *Mobility conditions*

In order to test the effects of mobility, a random waypoint model was used. An initial configuration was chosen as in the static case. In order to simulate mobility, a new destination was chosen for each node, along with a speed to move at. The destination was again chosen randomly using a uniform distribution. The speed was chosen in the interval $[0, V]$, again with uniform distribution.

Each time a node reached its destination, it entered a wait state. The wait period was randomly selected in the interval $[0, W]$. During this time, the node remained motionless at its current location. At the end of this period, a new destination and speed were selected, beginning a new mobility cycle.

Values for the five variables (S, N, R, V, and W) were chosen to be realistic and consistent with those used in previous protocol evaluations. For the lengths S and R, the values were chosen to be in typical units of metres (m). The wait times were chosen in units of seconds (s). Therefore, the speeds were in metres/second (m/s).

4. Results and analysis

In the following experiments, two aspects of MANETs were investigated. First, the overall connectivity of the network was explored. This included the properties of the nodes, their links and the links' lifetimes. Second, these links were put together to form routes within the network. Path lengths, lifetimes, the existence of multiple paths, and the importance of individual nodes were each addressed.

4.1. Network connectivity

One of the most fundamental requirements for a successful MANET is to have sufficient nodes in order to maintain connectivity across all nodes that want to be part of the network. There must be enough links in the network so that a node can reach any node that it desires. This is dependent on the actual positioning of the nodes, however the density of the nodes within the entire network determines the probability of the network becoming partitioned (so that one or more nodes become unreachable).

In the evaluation of a MANET, it is very important to consider whether or not all nodes are reachable. Disconnections, especially at the geographical edges of the network do exist in reality, and network protocols must be able to handle them. However, including these disconnections can seriously harm performance results if a node attempts to send to an unreachable node. If the network is severely partitioned, there may be little or no hope for creating a successful. Network parameters should be chosen to give the desired characteristics.

Although partitioning occurs due to the positions of the nodes, the probability of a partition occurring is dependent on the density of nodes within the network. Obviously, the higher the concentration of nodes within the network, the lower the likelihood that one or more of the nodes become isolated from the rest of the network. Fig. 7 shows the probabilities of the network becoming partitioned for different numbers of nodes within three different network areas. The transmission range used in the network will also affect partitioning.

Interestingly, the common configuration of 50 nodes in a 1000 m × 1000 m square area, with a transmission range of 250 m, gives a relatively high probability for partitioning. In this scenario, the partitioning is almost always one or two nodes being separated from the rest of the network (usually near the perimeter of the area). In a mobile scenario, there is a very high probability that

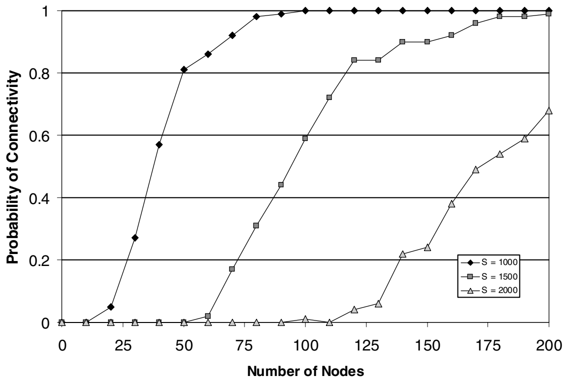

Fig. 7. Network connectivity. The probability of all nodes being reachable increases as the density of nodes increases.

certain nodes will become unreachable for some time during the simulation. If mobility is non-deterministic, then the only way to ensure that the network remains connected (with reasonable certainty) is to maintain a sufficiently high density of nodes.

4.2. Network density

Network density (or from graph theory, the degree of the nodes), has other effects within the network beyond basic network connectivity. A node communicates with its neighbours in order to send, receive, and forward traffic, but it must also compete with them for use of the wireless medium. A high density can cause contention problems, and reduce the efficiency of the channel usage. On the other hand, a high density can reduce the total number of transmissions required to broadcast messages throughout the network.

In a uniformly distributed network, a node's expected degree can be determined from the network size, the number of nodes, and the transmission range. In [16] this is given as

$$\text{Node degree} = (n-1)\frac{\text{area of transmission}}{\text{area of network}}, \quad (1)$$

where for this case, the area of transmission is the area of a circle of radius R, and the network area is the square of side S.

The same work [16] also describes the border effects of nodes near the boundary of the network. These nodes will have a lower expected degree, as their effective area of transmission is smaller, as part of their range lies outside the network. This will reduce the overall expected node degree from the value given by Eq. (1). In many experimental configurations, this reduction may be quite significant, if the border region (determined by the transmission range) is sizeable compared to the overall network. Fig. 8 shows this smaller node degree towards the edges of the network.

Figs. 9–11 show the effects of N, S, and R on the average number of neighbours. The results are as expected, slightly lower than predicted by Eq. (1), due to the network border effects. For example, for $N = 100$, $S = 1000$, $R = 250$, the expected node degree is approximately 19, while a value of

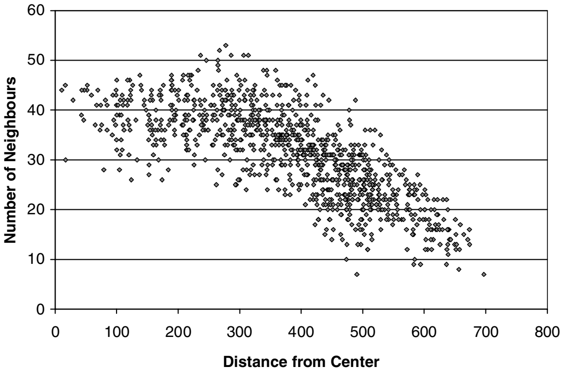

Fig. 8. Nodes in border region. In a network with borders, nodes near the border experience a decrease in the expected number of neighbours.

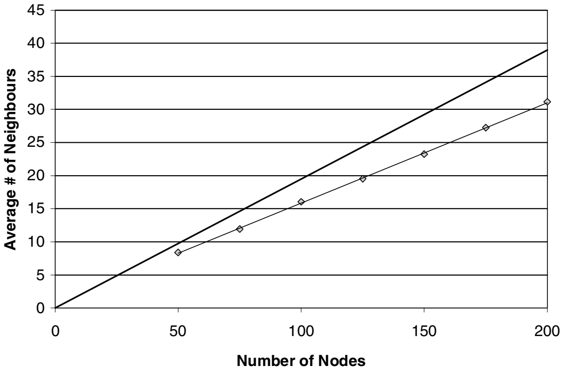

Fig. 9. The effects of increasing the number of nodes on the average number of neighbours. Experimentally obtained values are compared with the values given by Eq. (1).

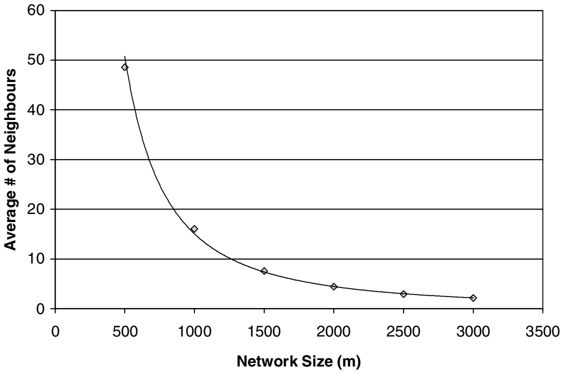

Fig. 10. The effects of increasing the size of the network on the average number of neighbours.

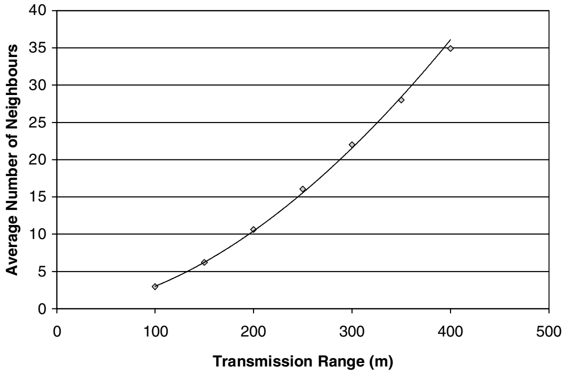

Fig. 11. The effects of increasing the transmission range on the average number of neighbours.

16 was obtained experimentally. As a result, node degree can reach sizeable levels very quickly, either by increasing the number of nodes per unit area, or especially by increasing the transmission range.

Fig. 12 shows cumulative distributions functions of number of neighbours for networks of 50, 100, and 200 nodes. Note that in some of these cases, certain nodes have links with over one quarter of the nodes in the network. In the configuration with 50 nodes, some nodes have reached as high as 17 neighbours (more than a third).

The possibility of very high numbers of neighbours stresses the need for an efficient method for sharing the available bandwidth. As any node must share the bandwidth with all of its neighbours, dividing the bandwidth that many ways greatly restricts the volume of traffic that can be sent by any particular node. If the MAC protocol is inefficient, particularly at high congestion levels, these conditions may be sufficient to reduce the overall effectiveness considerably.

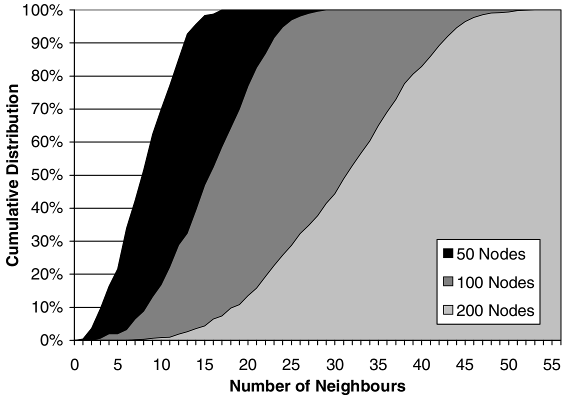

Fig. 12. The distribution of node degrees for networks with 50, 100, and 200 nodes.

The high level of connectivity also should be considered at the routing level. In on-demand protocols such as AODV and DSR, route discovery requires the flooding of the network in order to locate the destination. If flooding is accomplished by each node retransmitting a received discovery packet, a large degree of duplication occurs, with many nodes transmitting unnecessarily. This further emphasizes the need for systems such as OLSR's use of multipoint relays (MPRs) in order to minimize this duplication.

4.3. The random waypoint mobility model

To this point, the nodes in the network were distributed uniformly. However, this is not always a valid assumption. While the random waypoint model is often initialized with a uniform distribution of node locations, the distribution does not remain uniform. The resulting distribution can be seen in Fig. 13. This variation is due to the effect

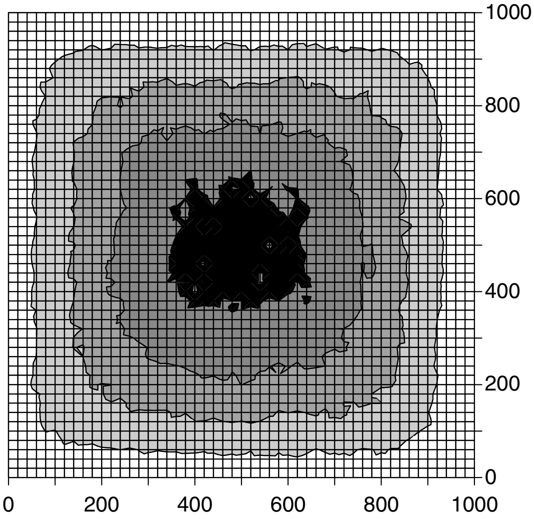

Fig. 13. The resulting probability distribution for the random waypoint mobility model. RWM results in a probability distribution heavily favouring the centre regions of the network.

of the borders on the movement of the nodes. As the destinations are selected using a uniform distribution, the nodes spend more time crossing through the centre of the network area.

The resulting distribution has a higher density near the centre of the network than a comparable uniform distribution. The edges have a lower density. The network has a higher overall density, due to the increased density at the centre. More nodes are affected by the increase in the central regions than by the corresponding reduction near the perimeter.

Several side effects should be considered. First, the border density is decreased; therefore, the probability of an edge node being disconnected from the network is increased. Second, the increased density at the centre of the network can greatly increase the node degree of all nodes in this region. This combination has the potential to make the scenario much more difficult than a similar uniform distribution would be. However, with a low overall node density, it may increase the likelihood of some nodes coming together to form at least a partial network.

This type of density distribution variation is a consequence of the mobility pattern. However reality is also unlikely to provide a truly uniform distribution. Ideally, the network should adapt to these variations. This could be used as an argument for variable-power transmissions, allowing the node to adapt its sending power in order to control the topology, minimize interference and maximize network throughput [9], and save energy [17].

Other mobility models are also sometimes used for MANET research. One particular example is the use of a "wrap-around" world, or toroidal environment, where each edge of a square environment is wrapped to meet with the opposite edge. Mobility in such a world is generated as a random direction, speed, and time, with the mobile "jumping" from one edge to the other. In such a world, the resulting distribution of nodes should remain uniform.

The distribution of nodes, whether uniform or not, must be considered when performing evaluations. One is not necessarily superior to the other, but one should be aware of the characteristics and consequences of their selection. While a uniform distribution may be easier to analyze or predict, in many situations a non-uniform distribution will be more realistic. In addition, the choice of a mobility model and network environment may have other consequences. For example, the wrap-around world will not exhibit any edge effects, which may be useful for simulating the center of a network, however ultimately it is unrealistic.

4.4. Maximum speeds

As the next section will show, the speeds chosen in the mobility model are critical to the characteristics of the network's links. Therefore, the choice of speeds (or maximum speed for the random waypoint model) should be made with care. To date, previous works have chosen maximum speeds ranging from 0 (static) to 30 m/s. Typically, these values have been chosen in order to test the routing protocol. Static networks are used to prove the correct operation of the protocol, while high maximum speeds result in frequent routing changes and stress the abilities of the protocol to rapidly react.

In fact, neither of these scenarios may be particularly realistic. The static case largely defeats the purpose of an ad hoc network. On the other hand, a 30 m/s maximum speed is equivalent to 108 km/h. Such a network, exhibiting random waypoint mobility motion at such a wide spread of speeds, is difficult to envision. It is certainly not a typical scenario. Instead, it should be considered an exceptionally difficult scenario, and is likely most useful only as a stress test in evaluating protocols under unmanageable conditions.

4.5. Link lifetimes

Once the nodes are mobile, links are created and broken as nodes move in and out of range of one another. Throughout the life of the network, links are formed and paths are built over these links. The stability of the links is of vital interest when constructing a path, or designing a routing protocol.

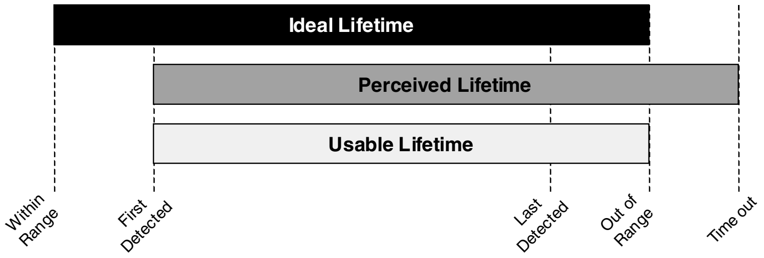

Fig. 14. Different methods for considering link lifetimes.

Several different times can be used in describing the life of a link, as illustrated in Fig. 14. First is the optimum lifetime of the link. This is the total time from when the nodes first move into range and a link could be created, until the link is broken by the nodes moving out of range. This value represents the maximum stable, usable period for the link. However, it is not actually the time that the link is available for use.

Therefore, a second time is also important. The usable lifetime of the link begins at the time when a node is first detected by another node, and lasts until the link is broken. The node may send to the known neighbour for the duration of this time, and assuming no interference (collisions), the packets should be properly received. Note, that the time of initial detection will certainly be different in opposite directions along the link. When a neighbour is detected, a node may then send to that neighbour, however, until it does the neighbour may not realize the reverse-relationship.

The link lifetime perceived by a node may in fact be different from the usable lifetime. Although the initial time, when the link is first detected, is known, the time of link breakage may not be. Instead, links are generally considered broken when the neighbour is not detected for a certain time period, or if it does not respond to attempts at direct communication (some number of RTS packets). At this time, the link is removed by the node, and it will no longer be used. However, this perceived link lifetime will in fact extend beyond the end of the existence of a usable link. Any packets transmitted during that time will not be successful, and are wasted effort.

One final time is also important. For communication that requires packet forwarding, the link must be included in a path before it is useful. For proactive routing protocols, this should be the first update period after the link is detected. However, for proactive protocols, a new route discovery must include the link in its path before the link is used. This discovery process may occur at any time during the link's lifetime, therefore the expected time to failure for the link, from the arbitrary time of route discovery, is half of the perceived link lifetime.

Although the density of nodes in the network affects the quantity of links formed, the lifetime is only dependent on the mobility model and transmission range used. In the random waypoint model, speeds and wait periods affect the link lifetimes. In Fig. 15 the average total (optimum) lifetime is

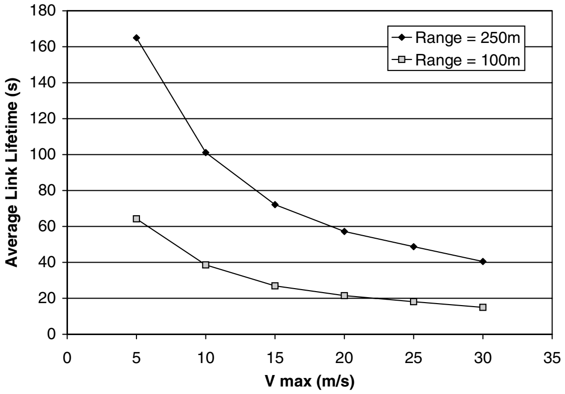

Fig. 15. The effect of the maximum speed parameter on average optimum link lifetimes using RWM.

shown as it varies against the maximum speed (S) for the nodes in that set of trials.

Increased node speed has a pronounced affect on the link lifetime. Although at low speeds (maximum speed of 5 m/s or 18 km/h) links last about 165 s on average (at 250 m transmission range,) at higher speeds links fail much sooner. When the maximum speed is set to 30 m/s (108 km/h), the lifetime falls to only 40 s. This means that a link in a newly formed route has an expected time to failure of only 20 s.

The link lifetimes for networks with shorter transmission ranges are even shorter. As nodes have smaller coverage areas, nodes do not have to move as far to move out of range. With total lifetimes ranging from 64 down to 15 s, these links tend to be very unstable if there is any mobility.

These are only average link lifetimes. In fact, the times are distributed as seen in Fig. 16. The distributions have long but light tails, as a few links last a very long time. However, the distributions have a significant weight stretching down to near-zero link lifetimes. This means that a large number of the links fail in a very short period of time.

Unfortunately, this is bad news for developing working MANETs. With a certain level of mobility, the majority of links become unusable. In fact, the length of time required to detect the fact that links have been formed or broken may exceed the lifetime of the link. Neighbours are detected when a packet is received from them. To facilitate this process, nodes will often broadcast HELLO packets at randomized intervals, in order to ensure that others will realize they are neighbours. Similarly, if nodes are not detected for a certain length of time (usually a multiple of the HELLO interval) then the link is considered to have broken. This interval must be sufficiently short in order to effectively maintain current link information, however shorter intervals greatly increase the amount of overhead added to the network.

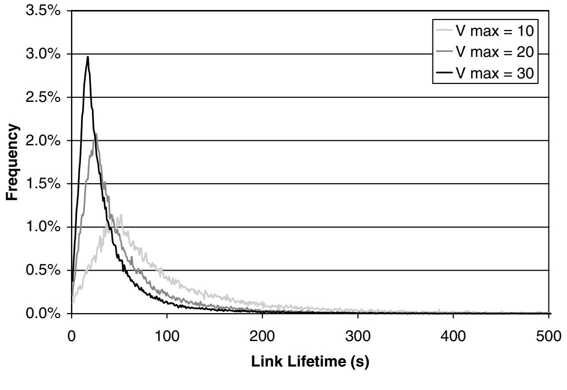

Fig. 16. Distributions of optimum link lifetimes.

4.6. Path lengths

Once links are created, paths can be formed. The length of the path is the number of links used along the path, or hop count. Most routing protocols strive to find the shortest paths, as this reduces the number of transmissions required in order to deliver packets to the destination, as well as minimizing the number of links that may break, causing a route failure.

As each wireless transmission has a maximum range, there is a strong relationship between the distance from sender to receiver and the minimum hop count. All routes must obey

$$\text{Shortest path} \geqslant \left\lceil \frac{\text{distance}}{\text{transmission range}} \right\rceil. \qquad (2)$$

The inequality depends on the density of nodes. The more nodes there are in the area, the more likely it is that there will be a path that is equal

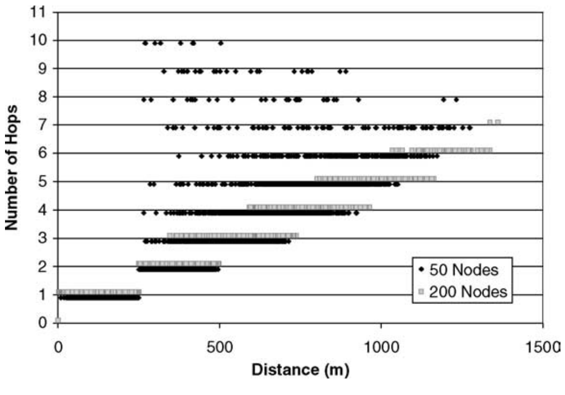

Fig. 17. Distribution of shortest path length vs. distance between the source and destination. An increase in the density of nodes in the network results in most paths being near optimal for number of hops vs. distance.

to the shortest possible path. In Fig. 17, the networks cover the same area, but there are four times as many nodes in the second network. While the first network has a large variation in the shortest path for any distance, the second network only rarely has to use anything but a minimum route, and even then it is never more than one additional hop.

For a dense network, this information could be useful. The distance between two nodes is a continuous function; it changes progressively with time. As the shortest path is closely tied to distance, it should only change when the distance crosses a threshold. Therefore, the hop count of the shortest path will only change in a stepwise manner, incrementing or decrementing by one each time.

The distribution of path lengths should also resemble the distribution of distances between nodes in the network. If the network is not sufficiently dense, some paths will be longer than the distance between source and destination would indicate. This skews the distribution slightly, towards higher hop counts. Increasing the number of nodes makes the distribution more normal, as seen in Fig. 18.

4.7. Multiple paths

Another consequence of the high node degree in a MANET is that alternate paths exist for many source–destination pairs. In many cases, these alternate paths are equal in length to the shortest path. Without considering an additional metric, these routes are interchangeable, however as they are constructed of a different set of links, the route may have a different failure time. If another metric is considered, often little or no penalty has to be paid in terms of hop count [18].

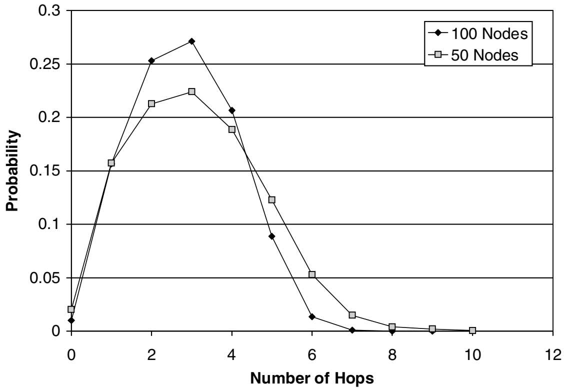

Fig. 18. Distribution of path lengths in two densities of networks. The denser network results in a tighter distribution of path lengths. The less dense network requires some paths to be much longer than optimal.

These multiple paths are not necessarily independent of each other. In a less dense MANET certain nodes are only reachable through certain other nodes. However, in a dense network, many nodes may be used interchangeably for routing purposes. Nodes are often close enough together that one node can freely substitute for another within a particular path. With a high node degree, adjacent nodes likely share many common neighbouring nodes. A node that is geographically closest to another node likely has the most common neighbours to that node. Therefore, for many paths (any path where the previous hop and next hop are part of the common neighbour set), one node can freely substitute for the other.

Fig. 19 shows the average number of paths found for each path length. These represent averaged values for networks with 50 nodes in a 1000 m^2 area, with a transmission range of 250 m. Obviously, for paths of hop count one there is only one possible path, however the average rises as high as 38 for paths of length 7, although there were relatively few actual source–destination pairs requiring this long of a path.

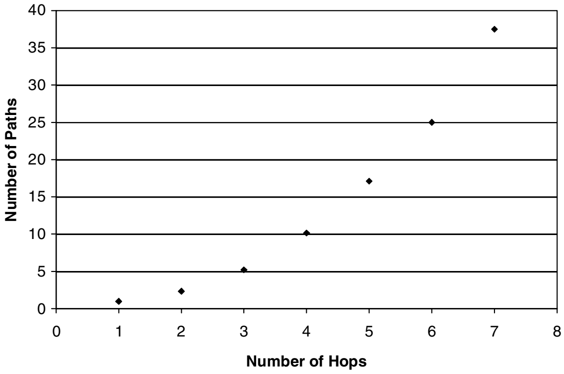

Fig. 19. Average number of paths available by hop count ($N = 50$, $S = 1000$, $R = 250$).

On average, there are fewer different paths for short paths than long ones. There are simply fewer nodes in the region that can take part in the routing process and still keep the hop count low. For longer routes, there are often a very large number of paths that could be exploited—in the case of the previous situation, certain source–destination pairs requiring a path of length 6 had over 200 different paths. Unfortunately, this is not true for every source–destination pair, at all times. Earlier, in Section 2.2.4, an "edge effect" was discussed, where the shortest path is a direct line of nodes, each right on the edge of the transmission range of the previous node. In this case, only a single path may exist, where no other route can be as short as that one. However, in this situation there are likely a large number of routes that are only one hop longer.

4.8. Path lifetime

With a large number of paths available for use, a decision must be made on which one should be used. Due to the overhead incurred when routing changes, ideally one would like to pick paths that will last the longest without breaking. Although other concerns may also figure into the decision, such as battery or processing resources, maintaining a working path is critical. If the path fails immediately after being created, then it is useless, and the effort is wasted.

In most networks it is not known when a path will fail. With no prior knowledge, any path is equally likely to be good as any other. However, a system may be able to benefit if it can efficiently take advantage of the last remaining path. Fig. 20 shows the distributions of path lifetimes for a high mobility network. Clearly the paths are extremely unstable. On average, half of the paths fail within about 4 s. Even in the best-case scenario, where the time of the last failing shortest path is used, half of the paths will fail within approximately 6 s. In the worst-case, if the worst possible shortest routes are chosen, half the nodes last less than 2 s.

How is this possible, when links (under high mobility) had a lifetime of about 40 s? First, the average time to failure of a link is half that, or 20 s. Next, many of the routes in the network have multiple links. The route fails whenever any link in the pathway breaks. Some of the links will fail before the expected time, and some of them after, but the first one to break causes the route failure.

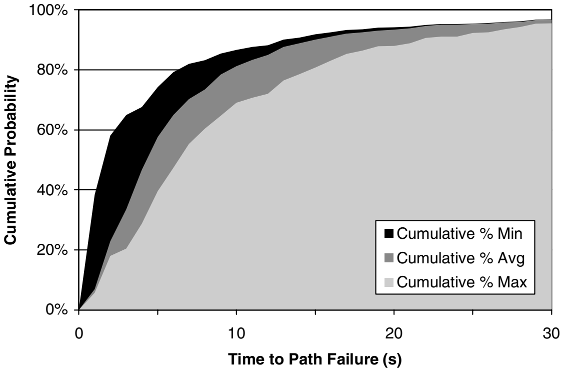

Fig. 20. Distribution of path lifetimes in a MANET ($V = [0–30]$, $W = [0]$). The cumulative probability of a path failing in a given time is given for the best path, the worst path, and an average of all paths.

5. Conclusions

In this paper, a number of statistics were collected from the topologies and mobility patterns of mobile ad hoc networks. Connectivity, node degrees, and path lengths were presented, along with link lifetimes and times to route failures. Overall, the results are cause for concern. Not only do many links break after a relatively short time period, but their short-lifespan is also propagated and exacerbated in the life spans of the routes.

The shortness of the route life spans is a problem. With route building already an expensive proposition in MANETs, these rapid routing changes are a severe challenge to the network. For today's protocols, the challenge is insurmountable. Current MANETs simply cannot effectively handle that level of change. Therefore, at least for know, system evaluations must take care to select a reasonable mobility setting, in order to achieve some reasonable level of performance.

The exceptionally high node degree that develops in ad hoc networks has the potential to be both the networks biggest asset, and its worst liability. This high degree creates a major contention issue for sharing the wireless medium. Scalability will be a significant problem, as adding nodes to a fixed size network can quickly cause the numbers of neighbours to skyrocket.

On the other hand, this node density could be one of the most distinguishing features of a MANET. The high node degree creates a multitude of routing options, unlike any other network. If this flexibility can be effectively harnessed, the routing protocol can be streamlined. The importance of any particular node to the routing process can be minimized, and resources can be better managed.

The characteristics investigated within this paper, and the statistics that were revealed, suggest that current MANET mechanisms, protocols, and technologies cannot handle many of the challenges it is possible to throw at them. Protocol development should continue, but must bear in mind the realistic limits that may be imposed by the inherent nature of the ad hoc environment. In order to conquer these limits, a revolution of ideas, rather than an evolution, may be necessary. For that to be achieved, the nature of those challenges must be better understood.

References

[1] C.E. Perkins, P. Bhagwat, Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers, Computer Communications Review 24 (4) (1994) 234–244.

[2] R.E. Bellman, Dynamic Programming, Princeton University Press, Princeton, 1957.

[3] C.-C. Chiang, Routing in clustered multihop, mobile wireless networks with fading channel, in: Proc. IEEE SICON'97, April 1997, pp. 197–211.

[4] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, L. Viennot, Optimized link state routing protocol for ad hoc networks, in: Proc. Int. Multitopic Conf. (IEEE INMIC), December 2001, pp. 62–68.

[5] C.E. Perkins, E.M. Royer, Ad-hoc on-demand distance vector routing, in: Proc. 2nd IEEE Workshop on Mobile Computing Systems and Applications, February 1999, pp. 90–100.

[6] D.B. Johnson, D.A. Maltz, Dynamic source routing in ad-hoc wireless networks, in: T. Imielinski, H. Korth (Eds.), Mobile Computing, Kluwer, Boston, 1996, pp. 153–181.

[7] M.R. Pearlman, Z.J. Haas, Determining the optimal configuration for the zone routing protocol, IEEE Journal on Selected Areas in Communications 17 (8) (1999) 1395–1414.

[8] S. Singh, M. Woo, C.S. Raghavendra, Power-aware routing in mobile ad hoc networks, in: Proc. ACM/IEEE Mobicom '98, October 1998, pp. 181–190.

[9] T.A. ElBatt, S.V. Krishnamurthy, D. Connors, S. Dao, Power management for throughput enhancement in wireless ad-hoc networks, in: Proc. Int. Conf. on Communications (IEEE ICC), June 2000, pp. 1503–1513.

[10] J.H. Chang, L. Tassiulas, Energy conserving routing in wireless as-hoc networks, in: Proc. IEEE Infocom 2000, March 2000, pp. 22–31.

[11] C.-K. Toh, Associativity-based routing for ad-hoc mobile networks, Wireless Personal Communications 4 (2) (1997) 103–139.

[12] R. Dube, C.D. Rais, K.-Y. Wang, S.K. Tripathi, Signal stability-based adaptive routing (SSA) for ad hoc mobile networks, IEEE Personal Communications 4 (1) (1997) 36–45.

[13] G. Lim, K. Shin, S. Lee, H. Yoon, J.S. Ma, Link stability and route lifetime in ad-hoc wireless networks, in: Proc. Int. Conf. on Parallel Processing Workshops (IEEE ICPPW'02), 2000.

[14] S. Chakrabarti, A. Mishra, QoS issues in ad hoc wireless networks, IEEE Communications Magazine 39 (2) (2001) 142–148.

[15] S. Xu, T. Saadawi, Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? IEEE Communications Magazine 39 (6) (2001) 130–137.

[16] H. Li, D. Yu, A statistical study of neighbour node properties in ad hoc network, in: Proc. Int. Conf. on Parallel Processing Workshops (IEEE ICPPW'02), 2002.

[17] S. Banerjee, A. Misra, Minimum energy paths for reliable communication in multi-hop wireless networks, in: Proc. ACM Mobihoc 2002, June 2002, pp. 146–156.

[18] R. Boutaba, Y. Iraqi, B. Ishibashi, Policy-based routing for ad hoc networks, in: Proc. Med-Hoc-Net 2002, August 2002.

Brent Ishibashi received the B.Sc. degree from the University of Guelph (Canada) in 2000. He recently completed his M.Math degree at the School of Computer Science of the University of Waterloo (Canada), and is now working towards a Ph.D. degree, also at the University of Waterloo. His research focuses on resource management in ad hoc network environments. Recent work has involved investigating the interactions between the link and network layers, particularly with finding approaches for better integrating medium access control with ad hoc routing.

Raouf Boutaba is currently an Associate Professor in the School of Computer Science of the University of Waterloo. Before that he was with the Department of Electrical and Computer Engineering of the University of Toronto. Before joining academia, he founded and was the director of the telecommunications and distributed systems division of the Computer Science Research Institute of Montreal (CRIM). He conducts research in the areas of network and distributed systems management and resource management in multimedia wired and wireless networks. He has published more than a hundred papers in refereed journals and conference proceedings. He is the recipient of the Premier's Research Excellence Award, a fellow of the Faculty of Mathematics of the University of Waterloo, and a distinguished lecturer of the IEEE Computer Society. He is the Chairman of the IFIP Working Group on Networks and Distributed Systems, the Vice Chair of the IEEE Communications Society Technical Committee on Information Infrastructure, and the Chair of the IEEE Communications Society Committee on Standards. He is on the advisory editorial board of JNSM, the editorial board of JCN, and the editorial board of Computer Networks.