An Upper Bound on Network Size in Mobile Ad-Hoc Networks

Michael Pascoe*, Javier Gomez, Victor Rangel
Department of Electrical Engineering
National Autonomous University of Mexico
Mexico City, Mexico
Email: (michael,javierg,victor)@fi-b.unam.mx

Miguel Lopez-Guerrero
Department of Electrical Engineering
Metropolitan Autonomous University
Mexico City, Mexico
Email: milo@xanum.uam.mx

Abstract—In this paper we propose a model to compute an upper bound for the maximum network size in mobile ad-hoc networks. Our model is based on the foundation that for a unicast route to be useful to initiate data transport, it is necessary that the time required to discover the route should be shorter than the time the route remains valid. From this model, we found that the node transmission range, mobility of nodes and number of contending nodes actually define the maximum feasible number of hops in a route, and therefore the maximum network size. Our model is derived from the combination of a route duration model, that we also derive in this paper, and a delay model for multi-hop routes extended from a single-hop delay model found in the literature. We evaluate our model numerically for different network conditions. To the best of our knowledge, this is the first time that a comparison between route discovery time and route duration is analyzed in order to establish the maximum network size in ad-hoc networks. We believe this is a fundamental scaling problem of ad-hoc networks that has not been looked at before from a mobility-delay perspective.

I. INTRODUCTION

A Mobile Ad-hoc Network (MANET) consists of a collection of mobile nodes connected by wireless links. In MANETs, nodes are free to move and organize without involving any infrastructure or centralized administration. Due to the limited transmission range of wireless radio transceivers, there may be a need for intermediate nodes, working as relays, to establish a communication path between source-destination pairs in the network. In Fig. 1 we show an arbitrary route from a source node $S$ to a destination node $D$ involving several forwarding nodes. Each circle in Fig. 1 represents the transmission range of each node, denoted by $R$. Node mobility causes frequent and unpredictable topology changes in the network. Routes, therefore, have limited lifetime.

Routing protocols for ad-hoc networks, can be classified into different categories according to the methods they use during the discovery and maintenance of routes. In proactive routing, routes from one node to all the other nodes in the network are discovered and maintained even if they are not needed. For reactive routing, nodes discover a route only when needed, usually by flooding the entire network with control packets. Although, reactive protocols usually exhibit higher latency compared to proactive protocols, because the former usually generate less signaling, they are preferably used in many practical scenarios. The model we present in this paper applies directly to reactive unicast routing protocols, e.g., DSR [1], AODV [2].

In general terms, reactive routing protocols are composed of two main mechanisms. Route Discovery is the mechanism by which a source node $S$, attempting to send data packets to a destination node $D$, discovers a route to node $D$. Route Maintenance is the mechanism by which nodes detect and locally repair any broken route that had been initially discovered and established by the route discovery mechanism. In case local route maintenance is not possible, node $S$ should attempt to discover another route to node $D$.

Any source-destination pair in MANETs should discover at least one valid route at the beginning of the first transmission. Such routes must guarantee that data transfer can take place, at least for a short period of time, even if both ends are located at the farthest opposite boundaries of the network. Based on this criterion, we can presuppose the existence of an upper limit on the maximum network size.

In order to determine an upper bound on the maximum network size in MANETs, from a mobility-delay perspective, we need to analyze how the route discovery process works in reactive routing protocols. The route discovery is composed of the following phases. When node $S$ attempts to send data packets to node $D$, it floods the network with control packets. The flooding begins when node $S$ broadcasts a route-
request packet. Neighbors of node $S$ receiving this packet will try to relay it once; this procedure continues until the entire network is flooded. During flooding, each relaying node experiences some channel contention and transmission delays. Let us denote such delay by $\Delta$ (see Fig. 1). In spite of these delays, the route-request packet will reach to node $D$ at a later time. When node $D$ receives the route-request packet, it sends a route-reply packet back to node $S$ using the same route; but in the opposite direction. A route from node $S$ to node $D$ will be established, from the source perspective, only when node $S$ receives the route-reply packet from node $D$. However, due to node mobility, there is a possibility that the route from node $D$ to node $S$ may fail before the route-reply packet reaches node $S$. This is a fundamental issue in the creation of routes for reactive routing protocols. This situation is illustrated in Fig. 1, where we observe that it takes some time for node $S$ to find node $D$ and also for node $D$ to reply back to node $S$. If by the time the reply is sent to node $S$, one of the intermediate nodes has already moved out from the route, the reply will not reach node $S$. In absence of a response, node $S$ will attempt to discover a route to node $D$ again. At this point, we can say that the operation of the routing protocol breaks down because no valid routes can be found to transport information.

In this paper, we show that the node transmission range, mobility of nodes and number of contending nodes actually define the maximum length of routes, measured by the number of intermediate hops, and therefore the maximum size of the network. Previous studies, related to scaling properties of ad-hoc networks, have mostly analyzed the traffic carrying capacity at the physical layer [3] [4] [5] and the MAC layer [6]. We argue that, for the traffic capacity of the network to be useful, we should have valid routes for a time interval that allows successful packet delivery between any source-destination pair. We believe this is a fundamental scaling problem in ad-hoc networks that has not been looked at before from a mobility-delay perspective.

Various authors have studied the performance of routing protocols in ad-hoc networks under different network conditions, such as the number of contending nodes, network size and mobility patterns. Most of these works were based on simulations [7] [8] [9]. However, they did not consider the existence of an upper bound on network size in mobile ad-hoc networks.

This paper makes a twofold contribution. First, we derive a route duration model that considers the delay involved during the route discovery phase. To the best of our knowledge, route duration analyses available in the literature, [10] [11] [12], assume that the route discovery time is negligible. This assumption is equivalent to consider that routes are discovered instantly. Second, we modify a transmission delay model for single-hop WLAN networks, found in [13], in order to obtain a delay model for multi-hop routes. By combining both models, we show that we are able to obtain a closed-form expression to compute the maximum length for valid routes in ad-hoc networks and therefore, the maximum network size.

The rest of the paper is organized as follows: Section II presents the route duration and route delay models. Section III presents an analysis to obtain the maximum length of routes in ad-hoc networks given the node transmission range, mobility and number of contending nodes in the network. Finally, in Section IV we present some conclusions.

II. MODEL COMPONENTS

In this section we summarize the route duration and route delay models required for the derivation of an upper bound for the maximum network size. First, we derive a route duration model in terms of the number of nodes involved in the route, node transmission range and speed of movement. This model also considers the route discovery time because, in order to obtain the maximum route length, we cannot ignore the time it takes to discover the route. Related details will be given below. Second, we derive a delay model for multi-hop routes based on a delay model for single-hop WLAN networks introduced in [13]. From the combination of both models we obtain an upper bound for the maximum network size, discussed later in Section III.

A. Route Duration Model

We define Route Duration, $T_{RD}$, as the interval measured from the instant when a valid route is discovered to the time when the route fails. This period of time specifies how long a route can be used to transport data. Now, we define Route Discovery Time, $T_D$, as the interval measured from the instant the source node sends the initial route request to the instant the route-reply packet reaches the destination node. Once the source node receives the route reply, a route has been established between the source-destination pair. Additionally, we define Route Failure Time, $T_F$, as the time measured from the instant the source node sends the initial route request to the instant the established route fails. We then formally define route duration as:

$$T_{RD} = \begin{cases} T_F - T_D & ; T_F \geq T_D \\ 0 & ; T_F < T_D \end{cases}$$  \hspace{1cm} (1)

In the previous definition, we consider that when an intermediate node, which might be a member of the route in the process of being discovered, leaves its overlapping region before the route is actually established, then there would be no route duration time for this hypothetical route.

A typical route is formed by a source node, a variable number of intermediate nodes and a destination node. The number of intermediate nodes depends on many factors, such as the distance between source and destination nodes, node transmission range and node density.

As illustrated in Fig. 2, let us consider a route formed by 3 mobile nodes, source node $S$, intermediate node $I$ and destination node $D$. In order for node $I$ to work as a forwarding node, it should be located within the intersection of the coverage zones of nodes $S$ and $D$ (overlapping region), represented by the shaded area in Fig. 2. Note that the size of the overlapping region depends on the distance between nodes $S$ and $D$. The time that node $I$ remains within this
Based on this analysis, we showed that the probability density function of routes with 3 mobile nodes, as the one shown in Fig. 2.

Even worse, it can make it impossible to discover a route. In time might considerably reduce the route duration time or, and trajectories of the overlapping regions and it also depends on the positions region can vary significantly because of the different sizes of intermediate nodes remain within their respective overlapping regions, see Fig. 1. In a route, formed by one or many intermediate nodes, the first intermediate node that abandons its overlapping region will cause a route failure.

In [14], we proposed a route duration model for ad-hoc networks in terms of the number of nodes involved in the route, node transmission range and speed of movement. In [14], we followed the commonly used assumption that routes are discovered instantly. This is valid only for scenarios where the route failure time ($T_F$) is orders of magnitude longer than the route discovery time ($T_D$). Otherwise, a long route discovery time might considerably reduce the route duration time or, even worse, it can make it impossible to discover a route. In [14], we performed exhaustive data analysis and simulations of routes with 3 mobile nodes, as the one shown in Fig. 2. Based on this analysis, we showed that the probability density function (PDF) of the route duration time $T$, for routes with 3 mobile nodes, is given by:

$$f_T(t) = \begin{cases} \sum_{j=1}^{2} \alpha_j e^{-(t - \beta_j)^2} & ; \quad t > 0 \\ 0 & ; \quad t < 0 \end{cases} \tag{2}$$

where, parameters $\alpha_j$, $\beta_j$, $\delta_j$, for $j = 1, 2$, were found by using a curve fitting method. The expression shown in (2) considered all possible initial positions and trajectories followed by the 3 mobile nodes ($S$, $I$ and $D$). More details are given in [14]. In Fig. 3a, we illustrate the PDF given by (2). We then analyzed routes formed by $N$ intermediate nodes as a superposition of $N$ 3-node routes (triplets). We found that the route duration time for a route formed by $N$ intermediate nodes can be obtained by determining the minimum of $N$ i.i.d. random variables defined by (2). In [14], we could not find a closed-form expression to compute the average route duration. We thus had to find it numerically, by repeatedly evaluating the route duration time for thousands of route sets formed by different number of intermediate nodes on each set and computing their average route duration.

Fig. 2. Route formed by 3 mobile nodes.

Fig. 3. (a) The PDF given by (2). (b) Transmission delays impact on route discovery and route duration.

The route duration model that we propose in this paper has two clear differences compared to [14]. First, we are able to derive a closed-form expression to compute the average route duration and second we take into consideration the route discovery time, which has been neglected in related models.

In order to find the maximum route length, which is a necessary step to find the upper bound on network size, we need to analyze the case when the route failure time is about the same order of magnitude as the route discovery time, $T_F \approx T_D$. After a careful inspection of Fig. 1, we observe that it takes some time for node $S$ to find node $D$, and also some time for node $D$ to reply back to node $S$. If we assume that each hop experiences practically the same average delay $\Delta$ in both ways, the average route discovery time ($T_D$), for routes formed by $N$ intermediate nodes, can be approximately found by:

$$T_D = 2(N + 1)\Delta, \tag{3}$$

where factor $2(N + 1)$ corresponds to the number of hops in a route formed by $N$ intermediate nodes, counted in both directions. In order to obtain the average delay $\Delta$, we propose to use a single-hop delay model found in [13]. This model will be referred later in this paper.

The derivation of the route duration model, presented in this paper, differs from other route duration models found in the literature because it considers hop by hop transmission delays in the calculation. In order to consider these delays in the analysis, we apply different time shifts to the PDF, given by (2). Each time shift corresponds to an individual transmission delay $\Delta$ experienced by each intermediate node during the route discovery process, see Fig. 3b. A time shift represents a time difference on the time of arrival of the route-request packet for each intermediate node. Time shifts applied to (2)
can be mathematically expressed by:

\[ f_{T_n}(t) = f_T(t - t_n) \quad (4) \]

where, \( t_n = n\Delta \) and \( \Delta \) is the individual delay that corresponds to a specific intermediate node, denoted by \( n \), \( n = 1, 2, \ldots, N \).

In Fig. 3b, we show a route formed by 3 intermediate nodes, which illustrates how route discovery and route duration are affected by transmission delays. The clocks describe the instant when the route request reaches each intermediate node. For instance, by the time the third intermediate node receives the route-request packet, the route duration associated to the first triplet has already consumed 2\( \Delta \) time units. We take this situation into consideration by shifting each PDF in time, as is also depicted in this figure.

The PDF associated to the new route duration model, would then be:

\[ f_{T_n}(t) = \left\{ \begin{array}{ll} \sum_{j=1}^{2} \alpha_j \beta_j \delta_j & ; \quad t \geq t_n \\ 0 & ; \quad t < t_n \end{array} \right. \quad (5) \]

where, parameters \( \alpha_j, \beta_j, \delta_j \), for \( j = 1, 2 \), can be found by using a curve fitting method, as used before for (2). \( T_n \) represents the time that a specific intermediate node remains within its overlapping region, \( T_n \geq 0 \).

Now, we need to derive a closed-form expression that allows us to compute the average route failure time (\( \bar{T}_F \)). By definition, the cumulative distribution function (CDF) associated to a PDF represents the probability that an intermediate node remains within its overlapping region a period of time within the interval \( T_n \leq t \). Let us denote such CDF by \( F_{T_n}(t) \).

In addition, the probability that an intermediate node remains within its overlapping region a time \( T_n > t \) would be given by the complementary cumulative distribution function (CCDF), i.e.,

\[ C_{T_n}(t) = P(T_n > t) = 1 - F_{T_n}(t). \quad (6) \]

We assume that the time each intermediate node remains within its respective overlapping region is an independent random variable. If the route is formed by \( N \) intermediate nodes, the probability that the route failure time (\( T_F \)) be within the interval \( T_F \leq t \), would be given by:

\[ P(T_F \leq t) = 1 - \prod_{n=1}^{N} P(T_n > t), \quad (7) \]

or

\[ P(T_F \leq t) = 1 - \prod_{n=1}^{N} C_{T_n}(t) = F_{T_F}(t), \quad (8) \]

where \( F_{T_F}(t) \) is the CDF associated to the failure time for a route formed by \( N \) intermediate nodes.

By definition, the route failure time is a positive random variable, so its average value \( \bar{T}_F \) could be found by using [15]:

\[ \bar{T}_F = \int_{0}^{\infty} (1 - F_{T_F}(\tau)) \, d\tau. \quad (9) \]

If we replace (8) in (9), we obtain:

\[ \bar{T}_F = \sum_{n=1}^{N} C_{T_n}(\tau) \, d\tau. \quad (10) \]

Apparently, the integral shown in (10) can only be solved by numerical methods for different values of \( N \). When solving (10) numerically, we have graphically observed that the average route duration time is inversely proportional to the number of intermediate nodes, \( N \), and speed of movement, \( v \). We use a curve fitting method to find a mathematical expression that represents the average failure time, in terms of \( N \) and \( v \). For this, we select an expression with two terms, because we have found experimentally that this expression represents the average failure time more accurately. An approximation of the average failure time, \( \bar{T}_F \), could thus be expressed as:

\[ \bar{T}_F = \frac{\kappa}{Nv} + \lambda(N + 1) \quad (11) \]

where parameters \( \kappa \) and \( \lambda \) can be found by using a curve fitting method.

Finally, if we replace (3) and (11) in (1), we can compute the average route duration time by means of:

\[ T_{RD} = \left\{ \begin{array}{ll} \frac{\kappa}{Nv} + \lambda(N + 1) - 2(N + 1)\Delta; & \bar{T}_F \geq T_D \\ 0 & ; \quad \bar{T}_F < T_D \end{array} \right. \quad (12) \]

B. Route Delay Model

**Round Trip Time (RTT)** is the time required for a packet to travel from a specific source node to a specific destination node and back again. This time depends on many factors including: the data transfer rate of the network, the number of intermediate nodes between source and destination nodes, the amount of traffic in the network, the MAC protocol, etc. The authors in [13] introduced a model to compute the average transmission delay, experienced by a packet transmitted by a node in the presence of \( c \) contending nodes in a saturated situation (i.e., all nodes always have packets to transmit) for single-hop WLAN networks. This model relies on the work presented by Bianchi in [16], which provides a model to evaluate the saturation throughput of the IEEE 802.11 MAC protocol under the hypothesis of ideal channel conditions (i.e., absence of hidden stations and transmission errors). In this paper, we focus on IEEE 802.11 MAC, because it has become a de facto standard in wireless ad-hoc networks. In case a different radio technology is used, a different delay model should be considered. The expression to compute the average transmission delay for single-hop routes, \( \bar{\Delta} \), is given by [13]:

\[ \bar{\Delta} = \bar{T}_b + \bar{T}_s, \quad (13) \]

where: \( \bar{T}_b \) is the average back-off time (i.e., contention time) and is given by \( \bar{T}_b = \alpha(W_{\min} - 1) + 1/\bar{q} \bar{T}_c \). Parameter \( \bar{T}_s \) is the average time that the channel is busy due to a successful transmission, given by \( \bar{T}_s = \bar{T}_{DIFS} + 3\bar{T}_{SIFS} + 4\bar{T}_e + \bar{T}_{RTS} + \bar{T}_{CTS} + \bar{T}_H + \bar{T}_p + \bar{T}_{ACK} \). Parameter \( \bar{T}_c \) is the time when a collision on the channel occurs, given
by $\overline{T}_c = T_{DIFS} + T_{RTS} + T_{\sigma}$. The terms $T_{DIFS}$ and $T_{SIFS}$ correspond to the inter-frame spaces used during the transmission. The terms $T_{RTS}$, $T_{CTS}$, $T_{H}$, $T_{P}$ and $T_{ACK}$ correspond to the time intervals allocated for the transmission of RTS, CTS, H (headers), P (payload or data) and ACK packets, respectively. $T_{\sigma}$ is the time when the channel is idle.

Additionally, $\alpha = (1 - P_{s})T_{c} + P_{s}T_{\overline{T}_s} + P_{t}(1 - P_{s})\overline{T}_c$ and $\beta = 2 - 2^{\alpha(q - 1)}$, where $q = 1 - p$ and $p$ is the conditional collision probability. $P_{s}$ is the probability that there is at least one transmission in the time slot. $P_{t}$ is the probability associated to a successful transmission on the channel. $W_{\min}$ is the minimum congestion window, $m$ is the maximum back-off stage. The authors in [13] found an approximation for the conditional collision probability $p$ in terms of $W_{\min}$ and $c$, i.e.,

$$p \approx \frac{2W_{\min}(c - 1)}{(W_{\min} + 1)^{2} + 2W_{\min}(c - 1)}.$$  \hspace{1cm} (14)

The probabilities $P_{s}$ and $P_{t}$, involved in this model, can be derived from the conditional collision probability $p$. Refer to [13] and [16] for more details.

Based on the transmission delay model for a single-hop route [13], described above, we propose to derive a delay model for a route formed by $N$ intermediate nodes. We define Route Delay as the time required for a packet to travel from node $S$ to node $D$ and back again, through a multi-hop route. If we assume that the route experiences practically the same average transmission delay $\Delta$ on each hop then, the average route delay $\overline{T}_{RTT}$ is proportional to the number of hops ($N + 1$), and can be computed by means of:

$$\overline{T}_{RTT} = 2(N + 1)\Delta.$$  \hspace{1cm} (15)

III. MAXIMUM NETWORK SIZE

As aforementioned, a route would be useful if, and only if, the route failure time is longer than the time interval required to discover the route. In Section II, we showed that route duration decreases with route length and the route delay increases with route length. The routes should therefore have a maximum length that meets both time conditions and guarantees a satisfactory communication between any pair of nodes of the network. The previous statements can be expressed analytically as:

$$\overline{T}_{RD} > \overline{T}_{RTT}.$$  \hspace{1cm} (16)

If we replace (12) and (15) in (16), we obtain:

$$\frac{\kappa}{Nv} + \lambda (N + 1) - 2(N + 1)\Delta > 2(N + 1)\Delta.$$  \hspace{1cm} (17)

Fig. 4 shows two sets of four curves each. The first set displays the average route duration time model versus number of intermediate nodes and the second set the average route delay versus number of intermediate nodes. In these curves, we consider two different values of contending nodes per transmission range area, i.e., $c = 10, 20$, and two different packet sizes, given by $P = 1500$ bytes (maximum IP packet size) and $P = 368$ bytes (average IP packet size) [17]. In these computations, we consider the node transmission range defined by the IEEE 802.11 standard (i.e., $R = 250 [m]$) and the speed of movement is $v = 1 [m/s]$.

From Fig. 4, we can infer that there is one intersection point on each pair of curves ($\overline{T}_{RD}$ and $\overline{T}_{RTT}$) with the same network conditions, i.e., contending nodes ($c$) and packet size ($P$). The abscissa of the intersection point corresponds to the maximum number of intermediate nodes, $N_{max}$, given the network conditions ($c$ and $P$). As long as $N \leq N_{max}$, it is guaranteed that useful routes can be discovered. When we equal both sides in (17) and solve the resulting equation for $N$, we obtain the maximum value $N_{max}$ given by:

$$N_{max} = \frac{1}{2}\left[-1 + \sqrt{1 + \frac{4\kappa}{v(4\Delta - \lambda)}}\right].$$  \hspace{1cm} (18)

where $[x]$ is the floor function of a real number $x$.

Fig. 5 shows a set of four curves displaying the maximum number of intermediate nodes, computed by means of (18) versus the speed of movement. Upon comparing these curves,
we observe that the maximum number of intermediate nodes is inversely proportional to the packet size and speed of nodes.

As above-mentioned, by limiting the maximum route length to a hop-count under the bound, given by (18), a communication path would be guaranteed for any source-destination pair in the network. So, if we assume that the maximum route length corresponds to the maximum diagonal of the network, we can easily compute the equivalent maximum network size. The maximum diagonal of the network, $D_{\text{max}}$, can be found by multiplying the mean distance between two adjacent nodes, $\bar{d}$, by $(N_{\text{max}} + 1)$, i.e.,

$$D_{\text{max}} = (N_{\text{max}} + 1) \bar{d}. \tag{19}$$

According to (18), the factor $(N_{\text{max}} + 1)$ corresponds to the maximum feasible number of hops in a route.

A simple method to obtain the mean distance between two adjacent nodes $\bar{d}$, found in (19), is to analyze a route with one intermediate node only, as the one shown in Fig. 2. If the distance between any source-destination pair, given by $d_{SD}$, is within the interval $R < d_{SD} < 2R$, then one intermediate node $I$, would be needed as a relay. If the distance between nodes $S$ and $D$ is uniformly distributed in the interval $R < d_{SD} < 2R$, its average value would be given by $\bar{d}_{SD} = (R + 2R)/2 = 1.5R$. Finally, the mean distance between either $S-I$ or $I-D$ corresponds to $\bar{d} = \bar{d}_{SD}/2 = 0.75R$. Other methods to find $\bar{d}$ would be to compute the average length of a MST (Minimum Spanning Tree) or by extensive network simulations.

IV. CONCLUSIONS

In this paper, we proposed a model to determine the upper bound on the maximum network size of wireless ad-hoc networks. The upper bound on the maximum network size is found by determining the maximum feasible number of intermediate nodes, $N_{\text{max}}$, in any route of the network. First, we approached this problem by proposing a new route duration model for routes formed by $N$ intermediate nodes that takes into account the average route discovery time. Based on this model, we provided an approximation to compute the average route failure time and, therefore, the average route duration time. Second, we used an average delay model for single-hop routes, found in the literature, to derive a route delay model for multi-hop routes. From both models, we obtained a closed-form expression to compute the maximum feasible number of intermediate nodes that guarantees a reliable communication path for any source-destination pair. Thus, the maximum network size can be estimated. Numerical calculations were developed to evaluate this study for different network conditions. From this analysis, we concluded that the maximum number of intermediate nodes is inversely proportional to the packet size and speed of nodes. This model can be used to scale up/down the network size as to meet minimum route duration requirements to guarantee a communication path for any source-destination pair in ad-hoc networks.

ACKNOWLEDGMENT

This work is supported in part by research funds from CONACyT grant 47197-Y and PAPIIT grant IN105307. * PhD student supported by a UNAM/DGEP scholarship.

REFERENCES