Performance Analysis of Geographic Routing Protocols in Ad Hoc Networks

Don Torrieri,* Salvatore Talarico,† and Matthew C. Valenti‡
*U.S. Army Research Laboratory, Adelphi, MD, USA
†West Virginia University, Morgantown, WV, USA

Abstract—Geographic routing protocols greatly reduce the requirements of topology storage and provide flexibility in the accommodation of the dynamic behavior of ad hoc networks. This paper presents performance evaluations and comparisons of two geographic routing protocols and the popular AODV protocol. The trade-offs among the average path reliabilities, average conditional delays, average conditional number of hops, and area spectral efficiencies and the effects of various parameters are illustrated for finite ad hoc networks with randomly placed mobiles. This paper uses a dual method of closed-form analysis and simple simulation that is applicable to most routing protocols and provides a much more realistic performance evaluation than has previously been possible. Some features included in the new analysis are shadowing, exclusion and guard zones, and distance-dependent fading.

I. INTRODUCTION

Mobile ad hoc networks often use the ad-hoc on-demand distance-vector (AODV) routing protocol [1], which discovers and maintains multihop paths between source mobiles and destination mobiles. However, these paths are susceptible to disruption due to changes in the fading, terrain, and interference, and hence the control overhead requirements are high. An alternative class of routing protocols that do not maintain established routes between mobiles are the geographic routing protocols. These protocols require only a limited amount of topology storage by mobiles and provide flexibility in the accommodation of the dynamic behavior of ad hoc networks [2], [3].

Among the many varieties of geographic routing protocols, two representative ones are evaluated in this paper: greedy forwarding, which uses beacons, and maximum progress routing, which is contention based. The trade-offs among the average path reliabilities, average conditional delays, average conditional number of hops, and area spectral efficiencies and the effects of various parameters are illustrated for finite ad hoc networks with randomly placed mobiles. A comparison is made with the AODV routing protocol to gain perspective about the advantages and disadvantages of geographic routing.

This paper uses a dual method of closed-form analysis and simple simulation to provide a realistic performance evaluation of the three routing protocols. The method performs spatial averaging over network realizations by exploiting the deterministic geometry of [4] rather than the conventional stochastic geometry [5], thereby eliminating many unrealistic restrictions and assumptions, as explained in [6]. The method has great generality and can be applied to the performance evaluation of most other routing protocols.

II. NETWORK MODEL

The network comprises \( M + 2 \) half-duplex mobiles in an arbitrary two- or three-dimensional region. The variable \( X_i \) represents both the \( i^{th} \) mobile and its location, and \( ||X_j - X_i|| \) is the distance from the \( i^{th} \) mobile to the \( j^{th} \) mobile. Mobile \( X_0 \) serves as the reference transmitter or message source, and mobile \( X_M+1 \) serves as the reference receiver or message destination. The other \( M \) mobiles \( X_1, \ldots, X_M \) are potentially relays or sources of interference. Each mobile uses a single omnidirectional antenna.

Exclusion zones surrounding the mobiles, which ensure a minimum physical separation between two mobiles, have radii set equal to \( r_{ex} \). The mobiles are uniformly distributed throughout the network area outside the exclusion zones, according to a uniform clustering model [7].

The mobiles of the network transmit asynchronous quadrature direct-sequence signals. For such a network, interference is reduced after despreading by the factor \( h / G \), where \( G \) is the processing gain or spreading factor, and \( h \) is the chip factor [4], which reduces interference due to its asynchronism. Let \( P_i \) denote the received power from \( X_i \) at the reference distance \( d_0 \) before despreading when fading and shadowing are absent. After the despreading, the power of \( X_i \)'s signal at the mobile \( X_j \) is

\[
\rho_{i,j} = \tilde{P}_i g_{i,j} 10^{\frac{\xi_{i,j}}{10}} f(||X_j - X_i||)
\]

where \( \tilde{P}_i = P_i \) for the desired signal, \( \tilde{P}_i = h P_i / G \) for an interferer, \( g_{i,j} \) is the power gain due to fading, \( \xi_{i,j} \) is a shadowing factor, and \( f(\cdot) \) is a path-loss function. The path-loss function is expressed as the power law

\[
f(d) = \left( \frac{d}{d_0} \right)^{-\alpha}, \quad d \geq d_0
\]

where \( \alpha \geq 2 \) is the path-loss exponent, \( d_0 \) is sufficiently far that the signals are in the far field, and \( r_{ex} \geq d_0 \).
The $\{g_{i,j}\}$ are independent with unit-mean but are not necessarily identically distributed; i.e., the channels from the different $\{X_i\}$ to $X_j$ may undergo fading with different distributions. For analytical tractability and close agreement with measured fading statistics, Nakagami fading is assumed, and $g_{i,j} = a_{i,j}^2$, where $a_{i,j}$ is Nakagami with parameter $m_{i,j}$. It is assumed that the $\{g_{i,j}\}$ remain fixed for the duration of a time interval but vary independently from interval to interval (block fading).

In the presence of shadowing with a lognormal distribution, the $\{\xi_{i,j}\}$ are independent zero-mean Gaussian random variables with variance $\sigma^2_i$. For ease of exposition, it is assumed that the shadowing variance is the same for the entire network, but the results may be easily generalized to allow for different shadowing variances over parts of the network. In the absence of shadowing, $\xi_{i,j} = 0$. While the fading may change from one transmission to the next, the shadowing remains fixed for the entire session.

The service probability $\mu_i$ is defined as the probability that mobile $X_i$ can serve as a relay along a path from a source to a destination, and $1 - \mu_i$ is the probability that $X_i$ is a potential interferer. A mobile may not be able to serve as a relay in a path from $X_0$ to $X_{M+1}$ because it is already receiving a transmission, is already serving as a relay in another path, is transmitting, or is otherwise unavailable.

With interference probability $p_i$, a potentially interfering $X_i$ transmits in the same time interval as the desired signal. The $\{p_i\}$ can be used to model the servicing of other streams, controlled silence, or failed link transmissions and the resulting retransmission attempts. Mobiles $X_0$ and $X_{M+1}$ do not cause interference. When the mobile $X_j$ serves as a potential relay, we set $p_j = 0$.

Let $N^0$ denote the noise power, and the indicator $I_i$ denote a Bernoulli random variable with probability $P[I_i = 1] = p_i$. Since the despreading does not significantly affect the desired-signal power, (1) and (2) imply that the instantaneous signal-to-interference-and-noise ratio (SINR) at the mobile $X_j$ for a desired signal from mobile $X_k$ is

$$\gamma_{k,j} = \frac{g_{k,j} \Omega_{k,j}}{\Gamma^{-1} + \sum_{i=1,i\neq k}^{M} I_i g_{i,j} \Omega_{i,j}}$$

(3)

where

$$\Omega_{i,j} = \begin{cases} 10^{\xi_{i,j}/10} ||X_j - X_k||^{-\alpha} & i = k \\ h P_i \frac{10^{\xi_{i,j}/10} ||X_j - X_k||^{-\alpha}}{GP_k} & i \neq k \end{cases}$$

(4)

is the normalized power of $X_i$ at $X_j$, and $\Gamma = d_0^2 P_k/N$ is the SNR when $X_k$ is at unit distance from $X_j$ and fading and shadowing are absent.

The outage probability quantifies the likelihood that the interference, shadowing, fading, and noise will be too severe for useful communications. Outage probability is defined with respect to an SINR threshold $\beta$, which represents the minimum SINR required for reliable reception. In general, the value of $\beta$ depends on the choice of coding and modulation. An outage occurs when the SINR falls below $\beta$. In [4], closed-form expressions are provided for the outage probability conditioned on the particular network geometry and shadowing factors. Let $\Omega_j = \{\Omega_{0,j}, ..., \Omega_{M+1,j}\}$ represent the set of normalized powers at $X_j$. Conditioning on $\Omega_j$, the outage probability of the link from $X_k$ to receiver $X_j$ is

$$\epsilon_{k,j} = P[\gamma_{k,j} \leq \beta \mid \Omega_j].$$

(5)

The conditioning enables the calculation of the outage probability for any specific network geometry, which cannot be done using tools based on stochastic geometry. The closed-form equations for $\epsilon_{k,j}$ are used in the subsequent performance evaluations of the routing protocols.

III. ROUTING MODELS

A. Routing Protocols

The three routing protocols that are considered are reactive or on-demand protocols that only seek routes when needed and do not require mobiles to store details about large portions of the network. The AODV protocol relies on flooding to seek the fewest-hops path during its path-discovery phase. The flooding diffuses request packets simultaneously over multiple routes for the purpose of discovering a successful route to the destination despite link failures along some potential paths. When the first request packet reaches the destination, backtracking by an acknowledgement packet establishes the route the request packet followed as the single static fewest-hops path for subsequent message packets during a message-delivery phase. Subsequent receptions of request packets by the destination are ignored. There is a high overhead cost in establishing the fewest-hops path during the path-discovery phase, and the fewest-hops path must be used for message delivery before changes in the channel conditions cause an outage of one or more of its links.

Geographic protocols limit information-sharing costs by minimizing the reliance of mobiles on topology information [2], [3]. Since geographic routing protocols make routing decisions on a hop-by-hop basis, they do not require a flooding process for path discovery. Two geographic routing protocols are examined: the greedy forwarding protocol and the maximum progress protocol. Both geographic routing protocols assume that each mobile knows its physical location and the direction towards the destination.

The greedy forwarding protocol relies on beacons, which are mobiles that periodically broadcast information about their locations. A source forwards a packet
to a relay that is selected from a set of neighboring beacons that are modeled as the set of active mobiles that lie within a transmission range of radius \( r_i \). The next link in the path from source \( X_0 \) to destination \( X_{M+1} \) is the link to the relay within the transmission range that shortens the remaining distance to \( X_{M+1} \) the most. There is no path-discovery phase because the relays have the geographic information necessary to route the messages to the destination.

The maximum progress protocol is a contention-based protocol that does not rely on beacons but comprises alternating path-discovery phases and message-delivery phases. During a path-discovery phase, a single link to a single relay is discovered. During the following message-delivery phase, a packet is sent to that relay, and then the alternating phases resume until the destination is reached. In a path-discovery phase, the next relay in a path to the destination is dynamically selected at each hop of each packet and depends on the local configuration of available relays. A source or relay broadcasts Request-to-Send (RTS) messages to neighboring mobiles that potentially might serve as the next relay along the path to the destination. The RTS message includes the location of the transmitting source or previous relay. Upon receiving the RTS, a neighboring mobile initiates a timer that has an expiration time proportional to the remaining distance to the destination. When the timer reaches its expiration time, the mobile sends a Clear-to-Send (CTS) message as an acknowledgement packet to the source or previous relay. The earliest arriving CTS message causes the source or previous relay to launch the message-delivery phase by sending message packets to the mobile that sent that CTS message, and all other candidate mobiles receiving that CTS message cease operation of their timers.

### B. Implementation of Path Selection

For the analysis and simulation, we draw a random realization of the network (topology) using the uniform clustering distribution of mobiles. The source and destination mobiles are placed, and then, one by one, the location of each remaining \( X_i \) is drawn according to a uniform distribution within the network region. However, if an \( X_i \) falls within the exclusion zone of a previously placed mobile, then it has a new random location assigned to it as many times as necessary until it falls outside all exclusion zones. Using the service probabilities, the set of potential relays is randomly selected for each simulation trial.

The routing protocols use a distance criterion to exclude a link from mobile \( X_i \) to mobile \( X_j \) as a link in one of the possible paths from \( X_0 \) to \( X_{M+1} \) if \( |X_j - X_{M+1}| > |X_i - X_{M+1}| \). These exclusions ensure that each possible path has links that always reduce the remaining distance to the destination. All links connected to mobiles that cannot serve as relays are excluded as links in possible paths from \( X_0 \) to \( X_{M+1} \). Links that have not been excluded are called eligible links.

The eligible links are used to determine the greedy-forwarding path from \( X_0 \) to \( X_{M+1} \) during its message-delivery phase. There is no path-discovery phase. If no path from \( X_0 \) to \( X_{M+1} \) can be found or if the message delivery fails, a routing failure is recorded.

A candidate link is an eligible link that does not experience an outage during the path-discovery phase. To identify the candidate links within each topology, we apply our analysis to determine the outage probability for each eligible link. A Monte Carlo simulation decides whether an eligible link is in an outage by sampling a Bernoulli random variable with the corresponding outage probability. A links that is not in an outage is called a candidate link.

For AODV, the candidate paths from \( X_0 \) to \( X_{M+1} \) are paths that can be formed by using candidate links. The candidate path with the fewest hops from \( X_0 \) to \( X_{M+1} \) is selected as the fewest-hops path. This path is determined by using the Dijkstra algorithm [8] with the unit cost of each candidate link. If two or more candidate paths have the fewest hops, the fewest-hops path is randomly selected from among them. If there is no set of candidate links that allow a path from \( X_0 \) to \( X_{M+1} \), then a routing failure occurs. If a fewest-hops path exists, then a Monte Carlo simulation is used to determine whether the acknowledgement packet traversing the path in the reverse direction is successful. If it is not or if the message delivery over the fewest-hops path fails, then a routing failure occurs.

A two-way candidate link is an eligible link that does not experience an outage in either the forward or the reverse direction during the path-discovery phase. A Monte Carlo simulation is used to determine the two-way candidate links. For the maximum progress protocol, the two-way candidate link starting with source \( X_0 \) with a terminating relay that minimizes the remaining distance to destination \( X_{M+1} \) is selected as the first link in the maximum-progress path. The link among the two-way candidate links that minimizes the remaining distance and is connected to the relay at the end of the previously selected link is added successively until the destination \( X_{M+1} \) is reached and hence the maximum-progress path has been determined. After each relay is selected, a message packet is sent in the forward direction to the selected relay. If no maximum-progress path from \( X_0 \) to \( X_{M+1} \) can be found or if a message delivery fails, a routing failure is recorded.

The CTS message transmitted by the maximum progress protocol during its path-discovery phase establishes guard zones [7]. Potentially interfering mobiles within the guard zones are silenced during the message-delivery phase of the maximum progress protocol. It is
assumed that the guard zones have sufficiently small radii $r_g$ that the CTS message is correctly decoded. Any potentially interfering mobile $X_i$ that lies in one of the guard zones surrounding the two mobiles at the ends of each link of a selected path is silenced by setting its $p_i = 0$ during message delivery.

C. Performance Metrics

Let $B$ denote the maximum number of transmission attempts over a link of the path. During the path-discovery phases, $B = 1$. During the message-delivery phases, $B \geq 1$ because message retransmissions over an established link are feasible. For each eligible or candidate link $l = (i, j)$, a Bernoulli random variable with failure probability $\epsilon_l$ is repeatedly drawn until there are either $B$ failures or success after $N_l$ transmission attempts, where $N_l \leq B$. The delay of link $l$ of the selected path is $N_l T + (N_l - 1) T_e$, where $T$ is the delay of a transmission over a link, and $T_e$ is the excess delay caused by a retransmission.

Each network topology $t$ is used in $K_t$ simulation trials. The path delay $T_{s,t}$ of a path from $X_0$ to $X_{M+1}$ for network topology $t$ and simulation trial $s$ is the sum of the link delays in the path during the message-delivery phase:

$$T_{s,t} = \sum_{l \in \mathcal{L}_{s,t}} [N_l T + (N_l - 1) T_e]. \tag{6}$$

where $\mathcal{L}_{s,t}$ is the set of links constituting the path. If there are $B$ transmission failures for any link of the selected path, then a routing failure occurs.

If there are $F_t$ routing failures for topology $t$ and $K_t$ simulation trials, then the probability of end-to-end success or path reliability within topology $t$ is

$$R_t = 1 - \frac{F_t}{K_t}. \tag{7}$$

Let $T_t$ denote the set of $K_t - F_t$ trials with no routing failures. If the selected path for trial $s$ has $h_{s,t}$ links or hops, then among the set $T_t$, the average conditional number of hops from $X_0$ to $X_{M+1}$ is

$$H_t = \frac{1}{K_t - F_t} \sum_{s \in T_t} h_{s,t}. \tag{8}$$

Let $T_d$ denote the link delay of packets during the path-discovery phase. The average conditional delay from $X_0$ to $X_{M+1}$ during the combined path-discovery and message-delivery phases is

$$D_t = \frac{1}{K_t - F_t} \sum_{s \in T_t} (T_{s,t} + 2c h_{s,t} T_d) \tag{9}$$

where $c = 0$ for the greedy forwarding protocol, and $c = 1$ for the maximum progress and AODV protocols. Let $A$ denote the network area and $\lambda = (M + 1)/A$ denote the density of the possible transmitters in the network. We define the normalized area spectral efficiency for the $K_t$ trials of topology $t$ as

$$A_t = \lambda \frac{1}{K_t} \sum_{s \in T_t} \frac{1}{T_{s,t} + 2c h_{s,t} T_d} \tag{10}$$

where the normalization is with respect to the bit rate or bits per channel use. The normalized area spectral efficiency is a measure of the end-to-end throughput in the network. After computing $R_t$, $D_t$, $H_t$, and $A_t$ for $T$ network topologies, we can average over the topologies to compute the topological averages: $\bar{R}$, $\bar{D}$, $\bar{H}$, and $\bar{A}$.

IV. Numerical Results

A host of network topologies and parameter values can be evaluated by the method described. Here, we consider a representative example that illustrates the trade-offs among the routing protocols. We consider a network occupying a circular region with normalized radius $r_{net} = 1$. The source mobile is placed at the origin, and the destination mobile is placed at distance $||X_{M+1} - X_0||$ from it. Times are normalized by setting $T = 1$. Each transmitted power $P_t$ is equal. There are no retransmissions during the path-discovery phases, whereas $B = 4$ during the message-delivery phases. A distance-dependent fading model is assumed, where a signal originating at mobile $X_i$ arrives at mobile $X_j$ with a Nakagami fading parameter $m_{i,j}$ that depends on the distance between the mobiles. We set

$$m_{i,j} = \begin{cases} 
3 & \text{if } ||X_j - X_i|| \leq r_l/2 \\
2 & \text{if } r_l/2 < ||X_j - X_i|| \leq r_l \\
1 & \text{if } ||X_j - X_i|| > r_l 
\end{cases} \tag{11}$$

where $r_l$ is the line-of-sight radius. The distance-dependent fading model characterizes the typical situation in which nearby mobiles most likely are in each other’s line-of-sight, while mobiles farther away from each other are not. Other fixed parameter values are $r_{ex} = 0.05$, $r_f = 0.2$, $T_e = 1.2$, $T_d = 1$, $M = 200$, $\beta = 0$ dB, $r_g = 0.15$, $K_t = 10^4$, $\Gamma = 0$ dB, $\alpha = 3.5$, and $\Upsilon = 2000$. The service and interference probabilities are assumed to have the same values for all mobiles so that $\mu_i = \mu$ and $p_i = p$. Unless otherwise stated, $G/h = 96$, $\alpha = 3.5$, $\mu = 0.4$, and $p = 0.3$. When shadowing is present, it has a lognormal distribution with $\sigma_s = 8$ dB. However, the transmitted packets encounter the same shadowing in both directions over the same link during both routing phases.

Figures 1 and 2 display the average path reliabilities of the request packets and acknowledgement packets, respectively, for the complete selected paths during the path-discovery phases of the AODV and maximum progress (MP) protocols. Figure 1 depicts the reliabilities both with and without shadowing as a function of the source-destination distance $||X_{M+1} - X_0||$. Shadowing is assumed in Fig. 2 and all subsequent figures. Figure 1 shows an initial decrease and then an increase in
average path reliability as the source-destination distance increases. This variation occurs because at short distances, there are very few relays that provide forward progress, and often the only eligible or candidate link is the direct link from source to destination. As the distance increases, there are more eligible and candidate links, and hence the network benefits from the diversity. Furthermore, as the destination approaches the edge of the network, the path benefits from a decrease in interference at the relays that are close to the destination. Figure 1 shows that during the request stage, the AODV protocol provides the better path reliability because it constructs several partial paths before the complete path is determined.

Since the relays are already determined in Fig. 2, the maximum progress protocol shows only a mild improvement with increasing source-destination distance, and this can be attributed almost entirely to the edge effect. It is observed in Fig. 2 that the AODV protocol has a relatively poor path reliability during the acknowledgement stage, which is due to the fact that a specified complete path must be traversed in the reverse direction, where the interference and fading may be much more severe. The maximum progress protocol does not encounter the same problem because the links in its paths are selected one-by-one with the elimination of links that do not provide acknowledgements. Although both the shadowing and the path-loss exponent $\alpha$ affect both the packets and the interference signals, the two figures indicate that the overall impact of more severe propagation conditions is detrimental for all distances.

Figure 3 displays the average path reliabilities for the message-delivery phases of the three protocols, assuming that the path-discovery phase, if used, has been successful. The figure illustrates the penalties incurred by the greedy forwarding (GF) protocol because of the absence of a path-discovery phase that eliminates links with excessive shadowing, interference, or fading and creates guard zones for the message-delivery phase.

The figure illustrates the role of the transmission range $r_t$ in determining average path reliability for greedy forwarding protocols. As $r_t$ increases, the links in the complete path are longer and less reliable. However, this disadvantage is counterbalanced by the increased number of potential relays and the reduction in the average number of links in a complete path.

Figure 4 shows the overall average path reliabilities for the combined path-discovery and message-delivery phases of all three routing protocols. The AODV protocol is the least reliable. The maximum progress protocol is much more reliable than the greedy forwarding protocol if $||X_{M+1} - X_0||$ is large, but is not as reliable if $||X_{M+1} - X_0|| < 0.35$ because of the relatively low reliability of its request packets.

The average conditional delay $\mathcal{D}$, the average conditional number of hops $\mathcal{H}$, and the normalized area spectral efficiency $\mathcal{A}$ for each routing protocol as a function of $||X_{M+1} - X_0||$ are displayed in Figures 5, 6, and 7, respectively. The greedy forwarding protocol has the highest $\mathcal{A}$ if $||X_{M+1} - X_0||$ is small, whereas the maximum progress protocol has the highest $\mathcal{A}$ if $||X_{M+1} - X_0||$ is large. The reason is the rapid loss of
reliability and increase in the average conditional delay of the greedy forwarding protocol when $|X_{M+1} - X_0|$ is large.

V. CONCLUSIONS

This paper presents performance evaluations and comparisons of two geographic routing protocols and the popular AODV protocol. The trade-offs among the average path reliabilities, average conditional delays, average conditional number of hops, and area spectral efficiencies and the effects of various parameters have been shown for a typical ad hoc network. Since acknowledgements are often lost due to the nonreciprocal interference and fading on the reverse paths, the AODV protocol has a relatively low path reliability, and its implementation is costly because it requires a flooding process. In terms of the examined performance measures, the greedy forwarding protocol is advantageous when the separation between the source and destination is small and the spreading factor is large, provided that the transmission range and the relay density are adequate. The maximum progress protocol is more resilient when the relay density is low and is advantageous when the separation between the source and destination is large.

The general methodology of this paper can be used to provide a significantly improved analysis of multihop routing protocols in ad hoc networks. Many unrealistic and improbable assumptions and restrictions of existing analyses can be discarded.

REFERENCES