Toward Reliable Data Delivery for Highly Dynamic Mobile Ad Hoc Networks

Shengbo Yang, Chai Kiat Yeo, and Bu Sung Lee

Abstract—This paper addresses the problem of delivering data packets for highly dynamic mobile ad hoc networks in a reliable and timely manner. Most existing ad hoc routing protocols are susceptible to node mobility, especially for large-scale networks. Driven by this issue, we propose an efficient Position-based Opportunistic Routing (POR) protocol which takes advantage of the stateless property of geographic routing and the broadcast nature of wireless medium. When a data packet is sent out, some of the neighbor nodes that have overheard the transmission will serve as forwarding candidates, and take turn to forward the packet if it is not relayed by the specific best forwarder within a certain period of time. By utilizing such in-the-air backup, communication is maintained without being interrupted. The additional latency incurred by local route recovery is greatly reduced and the duplicate relaying caused by packet reroute is also decreased. In the case of communication hole, a Virtual Destination-based Void Handling (VDVH) scheme is further proposed to work together with POR. Both theoretical analysis and simulation results show that POR achieves excellent performance even under high node mobility with acceptable overhead and the new void handling scheme also works well.

Index Terms—Geographic routing, opportunistic forwarding, reliable data delivery, void handling, mobile ad hoc network.

1 Introduction

Mobile ad hoc networks (MANETs) have gained a great deal of attention because of its significant advantages brought about by multihop, infrastructure-less transmission. However, due to the error prone wireless channel and the dynamic network topology, reliable data delivery in MANETs, especially in challenged environments with high mobility remains an issue. Traditional topology-based MANET routing protocols (e.g., DSDV, AODV, DSR [1]) are quite susceptible to node mobility. One of the main reasons is due to the predetermination of an end-to-end route before data transmission. Owing to the constantly and even fast changing network topology, it is very difficult to maintain a deterministic route. The discovery and recovery procedures are also time and energy consuming. Once the path breaks, data packets will get lost or be delayed for a long time until the reconstruction of the route, causing transmission interruption.

Geographic routing (GR) [2] uses location information to forward data packets, in a hop-by-hop routing fashion. Greedy forwarding is used to select next hop forwarder with the largest positive progress toward the destination while void handling mechanism is triggered to route around communication voids [3]. No end-to-end routes need to be maintained, leading to GR's high efficiency and scalability. However, GR is very sensitive to the inaccuracy of location information [4]. In the operation of greedy forwarding, the neighbor which is relatively far away from the sender is chosen as the next hop. If the node moves out of the sender's coverage area, the transmission will fail. In GPSR [5] (a very famous geographic routing protocol), the MAC-layer failure feedback is used to offer the packet another chance to reroute. However, our simulation reveals that it is still incapable of keeping up with the performance when node mobility increases.

In fact, due to the broadcast nature of the wireless medium, a single packet transmission will lead to multiple reception. If such transmission is used as backup, the robustness of the routing protocol can be significantly enhanced. The concept of such multicast-like routing strategy has already been demonstrated in opportunistic routing ([6], [7], [8]). However, most of them use link-state-style topology database to select and prioritize the forwarding candidates. In order to acquire the internode loss rates, periodic network-wide measurement is required, which is impractical for mobile environment. As mentioned in [9], the batching used in these protocols also tends to delay packets and is not preferred for many delay sensitive applications. Recently, location-aided opportunistic routing has been proposed [10] which directly uses location information to guide packet forwarding. However, just like the other opportunistic routing protocols, it is still designed for static mesh networks and focuses on network throughput while the robustness brought upon by opportunistic forwarding has not been well exploited.

In this paper, a novel Position-based Opportunistic Routing (POR) protocol is proposed, in which several forwarding candidates cache the packet that has been received using MAC interception. If the best forwarder does not forward the packet in certain time slots, suboptimal candidates will take turn to forward the packet according to a locally formed order. In this way, as long as one of the candidates succeeds in receiving and forwarding the packet, the data transmission will not be interrupted. Potential multipaths are exploited on the fly on a per-packet basis, leading to POR's excellent robustness.

• *The authors are with the Centre for Multimedia and Network Technology, School of Computer Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798.*
E-mail: {yang0201, asckyeo, ebslee}@ntu.edu.sg.

Manuscript received 25 Feb. 2010; revised 30 Nov. 2010; accepted 15 Dec. 2010; published online 17 Mar. 2011.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2010-02-0092.
Digital Object Identifier no. 10.1109/TMC.2011.55.

The main contributions of this paper can be summarized as follows:

- We propose a position-based opportunistic routing mechanism which can be deployed without complex modification to MAC protocol and achieve multiple reception without losing the benefit of collision avoidance provided by 802.11.
- The concept of in-the-air backup significantly enhances the robustness of the routing protocol and reduces the latency and duplicate forwarding caused by local route repair.
- In the case of communication hole, we propose a Virtual Destination-based Void Handling (VDVH) scheme in which the advantages of greedy forwarding (e.g., large progress per hop) and opportunistic routing can still be achieved while handling communication voids.
- We analyze the effect of node mobility on packet delivery and explain the improvement brought about by the participation of forwarding candidates.
- The overhead of POR with focus on buffer usage and bandwidth consumption due to forwarding candidates' duplicate relaying is also discussed. Through analysis, we conclude that due to the selection of forwarding area and the properly designed duplication limitation scheme, POR's performance gain can be achieved at little overhead cost.
- Finally, we evaluate the performance of POR through extensive simulations and verify that POR achieves excellent performance in the face of high node mobility while the overhead is acceptable.

The rest of this paper is organized as follows: we present the protocol design of POR and complementary mechanisms in Section 2. VDVH is depicted in Section 3. Section 4 analyzes the effect of node mobility on packet delivery and reveals the benefits brought about by the participation of forwarding candidates. Redundancy in POR, including memory consumption and duplicate relaying due to opportunistic forwarding will also be discussed. In Section 5, we evaluate the performance of proposed schemes by simulation and compare them with other routing protocols. Section 6 reviews the related work and conclusions are given in Section 7.

2 Position-Based Opportunistic Routing

2.1 Overview

The design of POR is based on geographic routing and opportunistic forwarding. The nodes are assumed to be aware of their own location and the positions of their direct neighbors. Neighborhood location information can be exchanged using one-hop beacon or piggyback in the data packet's header. While for the position of the destination, we assume that a location registration and lookup service which maps node addresses to locations is available just as in [5]. It could be realized using many kinds of location service ([11], [12]). In our scenario, some efficient and reliable way is also available. For example, the location of the destination could be transmitted by low bit rate but long range radios, which can be implemented as periodic beacon, as well as by replies when requested by the source.

When a source node wants to transmit a packet, it gets the location of the destination first and then attaches it to the packet header. Due to the destination node's movement, the multihop path may diverge from the true location of the final destination and a packet would be dropped even if it has already been delivered into the neighborhood of the destination. To deal with such issue, additional check for the destination node is introduced. At each hop, the node that forwards the packet will check its neighbor list to see whether the destination is within its transmission range. If yes, the packet will be directly forwarded to the destination, similar to the destination location prediction scheme described in [4]. By performing such identification check before greedy forwarding based on location information, the effect of the path divergence can be very much alleviated.

In conventional opportunistic forwarding, to have a packet received by multiple candidates, either IP broadcast or an integration of routing and MAC protocol is adopted. The former is susceptible to MAC collision because of the lack of collision avoidance support for broadcast packet in current 802.11, while the latter requires complex coordination and is not easy to be implemented. In POR, we use similar scheme as the MAC multicast mode described in [13]. The packet is transmitted as unicast (the best forwarder which makes the largest positive progress toward the destination is set as the next hop) in IP layer and multiple reception is achieved using MAC interception. The use of RTS/CTS/DATA/ACK significantly reduces the collision and all the nodes within the transmission range of the sender can eavesdrop on the packet successfully with higher probability due to medium reservation.

As the data packets are transmitted in a multicast-like form, each of them is identified with a unique tuple *(src_ip, seq_no)* where *src_ip* is the IP address of the source node and *seq_no* is the corresponding sequence number. Every node maintains a monotonically increasing sequence number, and an *ID_Cache* to record the ID *(src_ip, seq_no)* of the packets that have been recently received. If a packet with the same ID is received again, it will be discarded. Otherwise, it will be forwarded at once if the receiver is the next hop, or cached in a *Packet List* if it is received by a forwarding candidate, or dropped if the receiver is not specified. The packet in the *Packet List* will be sent out after waiting for a certain number of time slots or discarded if the same packet is received again during the waiting period (this implicitly means a better forwarder has already carried out the task).

The basic routing scenario of POR can be simply illustrated in Fig. 1. In normal situation without link break, the packet is forwarded by the next hop node (e.g., nodes A, E) and the forwarding candidates (e.g., nodes B, C; nodes F, G) will be suppressed (i.e., the same packet in the *Packet List* will be dropped) by the next hop node's transmission. In case node A fails to deliver the packet (e.g., node A has moved out and cannot receive the packet), node B, the forwarding candidate with the highest priority, will relay the packet and suppress the lower priority candidate's forwarding (e.g., node C) as well as node S. By using the feedback from MAC layer, node S will remove node A from the neighbor list and select a new next hop node for the

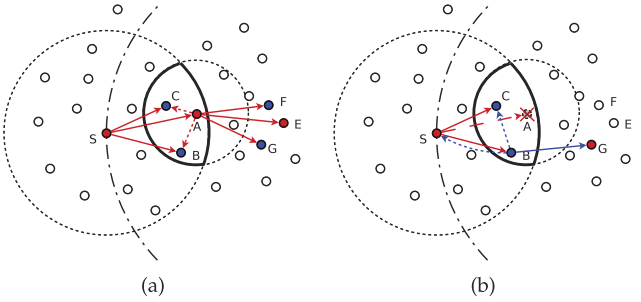


Fig. 1. (a) The operation of POR in normal situation. (b) The operation of POR when the next hop fails to receive the packet.

subsequent packets. The packets in the interface queue taking node A as the next hop will be given a second chance to reroute. For the packet pulled back from the MAC layer, it will not be rerouted as long as node S overhears node B's forwarding.

2.2 Selection and Prioritization of Forwarding Candidates

One of the key problems in POR is the selection and prioritization of forwarding candidates. Only the nodes located in the forwarding area [14] would get the chance to be backup nodes. The forwarding area is determined by the sender and the next hop node. A node located in the forwarding area satisfies the following two conditions: 1) it makes positive progress toward the destination; and 2) its distance to the next hop node should not exceed half of the transmission range of a wireless node (i.e., $R/2$) so that ideally all the forwarding candidates can hear from one another. In Fig. 1, the area enclosed by the bold curve is defined as the forwarding area. The nodes in this area, besides node A (i.e., nodes B, C), are potential candidates. According to the required number of backup nodes, some (maybe all) of them will be selected as forwarding candidates. The priority of a forwarding candidate is decided by its distance to the destination. The nearer it is to the destination, the higher priority it will get. When a node sends or forwards a packet, it selects the next hop forwarder as well as the forwarding candidates among its neighbors. The next hop and the candidate list comprise the forwarder list. Algorithm 1 shows the procedure to select and prioritize the forwarder list. The candidate list will be attached to the packet header and updated hop by hop. Only the nodes specified in the candidate list will act as forwarding candidates. The lower the index of the node in the candidate list, the higher priority it has.

Algorithm 1. Candidate Selection

ListN : Neighbor List
ListC : Candidate List, initialized as an empty list
N_D *: Destination Node*
base : Distance between current node and N_D

if $find(ListN, N_D)$ **then**
  $next\_hop \leftarrow N_D$
  return
end if
for $i \leftarrow 0$ *to* $length(ListN)$ **do**
  $ListN[i].dist \leftarrow dist(ListN[i], N_D)$
end for
$ListN.sort()$
$next\_hop \leftarrow ListN[0]$
for $i \leftarrow 1$ *to* $length(ListN)$ **do**
  if $dist(ListN[i], N_D) \geq base$ *or* $length(ListC) = N$ **then**
    break
  else if $dist(listN[i], listN[0]) < R/2$ **then**
    $ListC.add(ListN[i])$
  end if
end for

TABLE 1
Forwarding Table in POR

(src_ip, dst_ip)	next_hop	candidate_list
(N1, N11)	N4	N5, N6
(N2, N12)	N7	N8, N5
...	...	...

Every node maintains a forwarding table for the packets of each flow (identified as source-destination pair) that it has sent or forwarded. Before calculating a new forwarder list, it looks up the forwarding table, an example is illustrated in Table 1, to check if a valid item for that destination is still available. The forwarding table is constructed during data packet transmissions and its maintenance is much easier than a routing table. It can be seen as a trade-off between efficiency and scalability. As the establishment of the forwarding table only depends on local information, it takes much less time to be constructed. Therefore, we can set an expire time on the items maintained to keep the table relatively small. In other words, the table records only the current active flows, while in conventional protocols, a decrease in the route expire time would require far more resources to rebuild.

2.3 Limitation on Possible Duplicate Relaying

Due to collision and nodes' movement, some forwarding candidates may fail to receive the packet forwarded by the next hop node or higher priority candidate, so that a certain amount of duplicate relaying would occur. If the forwarding candidate adopts the same forwarding scenario as the next hop node, which means it also calculates a candidate list, then in the worst case, the propagation area of a packet will cover the entire circle comprising the destination as the center and the radius can be as large as the distance between the source and the destination. To limit such duplicate relaying, only the packet that has been forwarded by the source and the next hop node is transmitted in an opportunistic fashion and is allowed to be cached by multiple candidates. In other words, only the source and the next hop node need to calculate the candidate list, while for the packet relayed by a forwarding candidate, the candidate list is empty. Actually, such scheme has already been implied in Fig. 1b (e.g., node B only forwards the packet to node G). In this way, the propagation area of a packet is limited to a certain band between the source and the destination, as illustrated in Fig. 2. Moreover, with the use of ID cache, duplicate packets will be dropped soon and would not propagate any further.

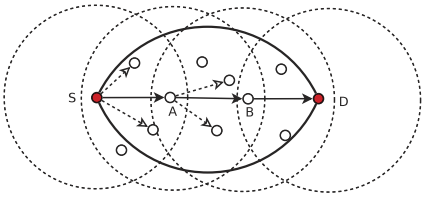


Fig. 2. Duplicate relaying is limited in the region enclosed by the bold curve.

2.4 MAC Modification and Complementary Techniques

2.4.1 MAC Interception

We leverage on the broadcast nature of 802.11 MAC: all nodes within the coverage of the sender would receive the signal. However, its RTS/CTS/DATA/ACK mechanism is only designed for unicast. It simply sends out data for all broadcast packets with CSMA. Therefore, packet loss due to collisions would dominate the performance of multicast-like routing protocols. Here, we did some alteration on the packet transmission scenario. In the network layer, we just send the packet via unicast, to the best node which is elected by greedy forwarding as the next hop. In this way, we make full utilization of the collision avoidance supported by 802.11 MAC. While on the receiver side, we do some modification of the MAC-layer address filter: even when the data packet's next hop is not the receiver, it is also delivered to the upper layer but with some hint set in the packet header indicating that this packet is overheard. It is then further processed by POR. Hence, the benefit of both broadcast and unicast (MAC support) can be achieved.

2.4.2 MAC Callback

When the MAC layer fails to forward a packet, the function implemented in POR—*mac_callback* will be executed. The item in the forwarding table corresponding to that destination will be deleted and the next hop node in the neighbor list will also be removed. If the transmission of the same packet by a forwarding candidate is overheard, then the packet will be dropped without reforwarding again; otherwise, it will be given a second chance to reroute. The packets with the same next hop in the interface queue which is located between the routing layer and MAC layer will also be pulled back for rerouting. As the location information of the neighbors is updated periodically, some items might become obsolete very quickly especially for nodes with high mobility. This scheme introduces a timely update which enables more packets to be delivered.

2.4.3 Interface Queue Inspection

One of the key points of POR is that when an intermediate node receives a packet with the same ID (i.e., same source address and sequence number), it means a better forwarder has already taken over the function. Hence, it will drop that packet from its packet list. Besides maintaining the packet list, we also check the interface queue. We do this because when the packet arrives at the routing layer, the same packet might have already been sent down to the lower layers by the current node. With additional inspection of the interface queue, we further decrease the duplicate packets appearing in the wireless channel.

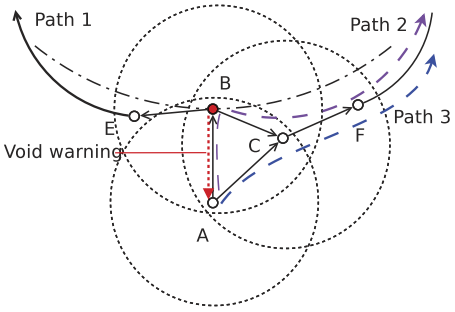


Fig. 3. Potential paths around the void.

3 Virtual Destination-Based Void Handling

In order to enhance the robustness of POR in the network where nodes are not uniformly distributed and large holes may exist, a complementary void handling mechanism based on *virtual destination* is proposed.

3.1 Trigger Node

The first question is at which node should packet forwarding switch from greedy mode to void handling mode. In many existing geographic routing protocols, the mode change happens at the void node, e.g., Node B in Fig. 3. Then, Path 1 (A-B-E-$\cdots$) and/or Path 2 (A-B-C-F-$\cdots$) (in some cases, only Path 1 is available if Node C is outside Node B's transmission range) can be used to route around the communication hole. From Fig. 3, it is obvious that Path 3 (A-C-F-$\cdots$) is better than Path 2. If the mode switch is done at Node A, Path 3 will be tried instead of Path 2 while Path 1 still gets the chance to be used. A message called void warning, which is actually the data packet returned from Node B to Node A with some flag set in the packet header, is introduced to trigger the void handling mode. As soon as the void warning is received, Node A (referred to as *trigger node*) will switch the packet delivery from greedy mode to void handling mode and rechoose better next hops to forward the packet. Of course, if the void node happens to be the source node, packet forwarding mode will be set as void handling at that node without other choice (i.e., in this case, the source node is the trigger node).

3.2 Virtual Destination

To handle communication voids, almost all existing mechanisms try to find a route around. During the void handling process, the advantage of greedy forwarding cannot be achieved as the path that is used to go around the hole is usually not optimal (e.g., with more hops compared to the possible optimal path). More importantly, the robustness of multicast-style routing cannot be exploited. In order to enable opportunistic forwarding in void handling, which means even in dealing with voids, we can still transmit the packet in an opportunistic routing like fashion, virtual destination is introduced, as the temporary target that the packets are forwarded to.

Virtual destinations are located at the circumference with the trigger node as center (Fig. 4), but the radius of the circle is set as a value that is large enough (e.g., the network

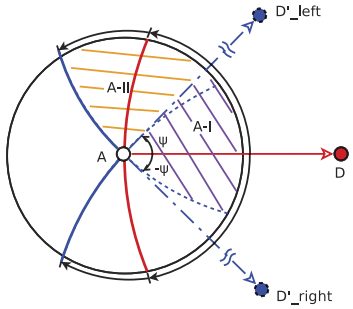


Fig. 4. Potential forwarding area is extended with virtual destination.

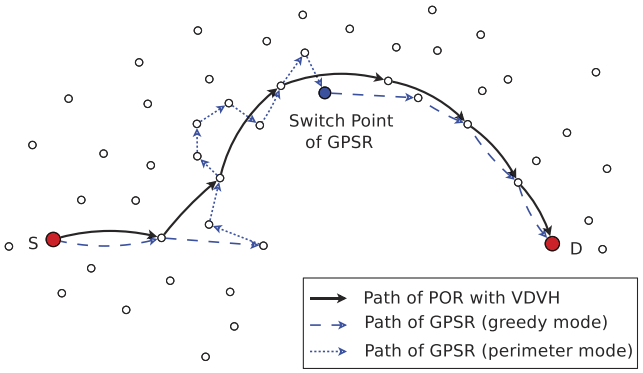


Fig. 5. The paths exploited by VDVH and GPSR.

diameter). They are used to guide the direction of packet delivery during void handling. Compared to the real destination D, a virtual destination (e.g., D'_{left} and D'_{right}) has a certain degree of offset, e.g., $\pm\psi$ ($\pi/4$ in our simulation) in Fig. 4. With the help of the virtual destination, the potential forwarding area is significantly extended. Strictly speaking, our mechanism cannot handle all kinds of communication voids, since not all the neighbors of the current node are covered. However, for most situations, it is effective. For those communication holes with very strange shape, a reposition scheme has been proposed [15] to smooth the edge of the hole. Given the work that has been done in [15], VDVH thus still has the potential to deal with all kinds of communication voids. Fig. 5 shows an example in which VDVH achieves the optimal path of seven hops while GPSR undergoes a much longer route of 15 hops.

3.3 Switch Back to Greedy Forwarding

A fundamental issue in void handling is when and how to switch back to normal greedy forwarding. From Fig. 4 we can see that the forwarding area in void handling can be divided into two parts: A-I and A-II. To prevent the packet from deviating too far from the right direction or even missing the chance to switch back to normal greedy forwarding, the candidates in A-I should be preferred and are thus assigned with a higher priority in relaying. Therefore, a scaling parameter is introduced for the candidates located in A-II. The progress toward the virtual destination made by these nodes is multiplied by a coefficient η $(0 < \eta < 1)$, called scaling parameter which is set as 0.75 in our experiment.

After a packet has been forwarded to route around the communication void for more than two hops (including two hops), the forwarder will check whether there is any potential candidate that is able to switch back. If yes, that node will be selected as the next hop, but the mode is still void handling. Only if the receiver finds that its own location is nearer to the real destination than the void node and it gets at least one neighbor that makes positive progress towards the real destination, it will change the forwarding mode back to normal greedy forwarding.

3.4 Path Acknowledgment and Disrupt Message

In VDVH, if a trigger node finds that there are forwarding candidates in both directions, the data flow will be split into two where the two directions will be tried simultaneously for a possible route around the communication void. In order to reduce unnecessary duplication, two control messages are introduced, namely, path acknowledgment and reverse suppression. If a forwarding candidate receives a packet that is being delivered or has been delivered in void handling mode, it will record a reverse entry. Once the packet reaches the destination, a path acknowledgment will be sent along the reverse path to inform the trigger node. Then, the trigger node will give up trying the other direction. For the same flow, the path acknowledgment will be periodically sent (not on per-packet basis; otherwise, there will be too many control messages). If there is another trigger node upstream, the path acknowledgment will be further delivered to that node, and so on.

On the other hand, if a packet that is forwarded in void handling mode cannot go any further or the number of hops traversed exceeds a certain threshold but it is still being delivered in void handling mode, a DISRUPT control packet will be sent back to the trigger node as reverse suppression. Once the trigger node receives the message, it will stop trying that direction.

4 Analysis

In this section, theoretical analysis on the robustness of POR will be conducted. The overhead inclusive of memory consumption and duplicate relaying will also be discussed. Since our focus lies on the effect of node mobility, an ideal wireless channel is assumed in the following part and the unit disc graph model will be used by default: a link between two nodes exists if and only if the distance between them is less than a certain threshold. When two nodes are located inside each others' coverage range (R), bidirectional data transmission between them can be achieved without failure.[1]

4.1 Robustness versus Mobility

Owing to node mobility, it is impossible that the location information of a node's neighbors which is maintained through beacon exchange is always up to date. Therefore, an error disc $b(\boldsymbol{x}, r_e)$ corresponding to each neighbor exists from the current node's perspective, with $\boldsymbol{x}$ as the latest obtained coordinate of the neighbor. The radius of the error disc r_e is the maximum deviation from $\boldsymbol{x}$ and the value of r_e varies with the elapsed time, t, since the last update and is defined as follows:

1. Here, we only consider the forwarding failure caused by node mobility while the effect of the unreliable wireless link has not be taken into consideration. We believe that in the light traffic case with retransmission scheme implemented in the MAC layer, node mobility should be the main factor resulting in the packet forwarding failure.

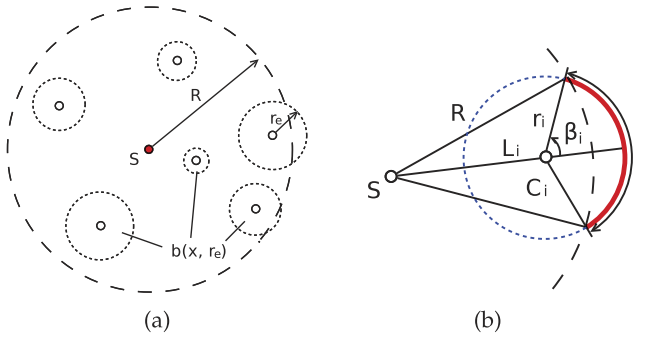


Fig. 6. (a) Network model. (b) Out of range caused by node's movement.

$$r_e(t) = v_r^m \cdot t, \quad 0 \le t < \delta t, \tag{1}$$

where v_r^m represents the maximum relative speed and δt is the neighbor update interval.[2] Since each node periodically updates its own location via beacon messages in an asynchronous fashion, the sampling error discs corresponding to different neighbors are of different sizes, as shown in Fig. 6.

Assume there are N forwarding candidates. We denote C_0 as the next hop and $C_{i,i\in[1,N]}$ as the forwarding candidates, so $C_{i,i\in[0,N]}$ is generally referred to as a forwarder (inclusive of the next hop and the forwarding candidates). Consider a particular forwarder C_i as an example. At a specific elapsed time t since the last update, given the values of the nodes' relative speeds $V_r = v$, the deviation from the latest updated location of C_i can be expressed as follows:

$$r_i(t, v) = v \cdot t, \quad 0 \le t < \delta t. \tag{2}$$

Further, given the distance between C_i and S, $L_i = l (0 < l \le R)$, we can see from Fig. 6b that only if $r_i(t, v) + l > R$, then C_i may move out of S's transmission range, and the corresponding probability is given by

$$\begin{aligned} p_i(t, v, l) &= P\{C_i \ is \ out \mid t, V_r = v, L_i = l\} \\ &= \begin{cases} \beta_i(t, v, l)/\pi, & vt + l > R, \\ 0, & otherwise, \end{cases} \end{aligned} \tag{3}$$

where $2\beta_i(t, v, l)$ represents the range of direction toward which the movement of C_i will cause it to be out of S's transmission range as shown with the bold curve in Fig. 6b. $\beta_i(t, v, l)$ can be expressed as

$$\beta_i(t, v, l) = arccos\left(\frac{R^2 - l^2 - (vt)^2}{2l \cdot vt}\right). \tag{4}$$

From (4), we can see that β_i grows as the value of vt increases, leading to a higher probability of C_i moving out of S's range. Since the mobility of a node is independent of each other, the relative speed between two nodes would not exceed $2v_{max}$ if the absolute node speed is distributed in $[v_{min}, v_{max}]$. On the other hand, t is upper bounded by δt. Thus, we can argue that node mobility and the location update interval are two main factors leading to packet forwarding failure due to the receiver moving, which will be evaluated at the end of this section.

2. Note that all the beacon messages will be correctly received by neighboring nodes due to the assumption of the ideal wireless channel. Therefore, δt here equals to the beacon broadcast interval.

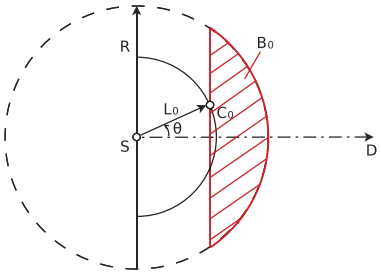


Fig. 7. L_0 with respect to C_0's location.

Without loss of generality, packet forwarding time from S to C_i is uniformly distributed in $[0, \delta t)$. We can thus give the average probability that C_i has moved out of S's transmission range when a packet from S is being sent to C_i as

$$p_i = \int_0^R \int_0^{2v_{max}} \int_0^{\delta t} p_i(t, v, l) \frac{1}{\delta t} dt f_{V_r}(v) dv f_{L_i}(l) dl, \tag{5}$$

where $f_{V_r}(v)$ and $f_{L_i}(l)$ are the pdf of V_r and L_i, respectively.

$f_{V_r}(v)$ can be evaluated numerically when the pdf of the absolute speed $f_V(v)$ is given [16]. For random way point (RWP) without pausing which we will use in our simulation, $f_V(v)$ is given by [17]

$$f_V(v) = \begin{cases} [v \cdot ln(v_{max}/v_{min})]^{-1}, & v_{min} \le v \le v_{max}, \\ 0, & otherwise. \end{cases} \tag{6}$$

The corresponding $f_{V_r}(v)$ is (refer to [18, (4.4)])

$$f_{V_r}(v) = \frac{1}{ln^2\left(\frac{v_{max}}{v_{min}}\right)} \int\int \frac{f_{V_r}(v|V_1 = v_1; V_2 = v_2)}{v_1 v_2} dv_1 dv_2, \tag{7}$$

where $f_{V_r}(v|V_1 = v_1; V_2 = v_2)$ is expressed as (refer to [18, (4.3)])

$$f_{V_r}(v|V_1 = v_1; V_2 = v_2) = \begin{cases} \dfrac{v}{\pi v_1 v_2 \sqrt{1 - \left(\frac{v_1^2 + v_2^2 - v^2}{2v_1 v_2}\right)^2}}, & if \ |v_1 - v_2| < v < v_1 + v_2, \\ 0, & otherwise. \end{cases} \tag{8}$$

For $f_{L_i}(l)$, the situation gets a bit more complex, as the location of $C_{i,i\in[1,N]}$ is closely associated with C_0.[3] Without loss of generality, we get $f_{L_0}(l)$ first and then calculate p_0. In most cases, $p_{i,i\in[1,N]}$ is smaller than p_0 since C_0 is usually nearer to the boundary of the sender's transmission range. By assuming $p_{i,i\in[1,N]} = p_0$, we can get some approximate result under challenging situations.

To calculate $f_{L_0}(l)$, we find the corresponding cumulative distribution function (CDF) $F_{L_0}(l)$ first. For C_0 to be the node with the highest priority, there must be no other nodes in the progress area toward the destination. As shown in Fig. 7, there is an arc of distance L_0 where C_0 has progressed towards the destination. The shaded area B_0 should not include any nodes (as required by the greedy routing protocol). Considering all possible positions on the arc

3. Recall the forwarding candidate selection scheme in Section 2.2. The forwarding area is partially determined by C_0, as only the nodes within half of the transmission range of C_0 will get the opportunity to become a forwarding candidate. That is why we say the location of $C_{i,i\in[1,N]}$ is closely associated with C_0, though nodes are independently located.

(defined as angular deviation), we integrate over the angular range and get $F_{L_0}(l)$ as follows:

$$F_{L_0}(l) = \int_{-\pi/2}^{\pi/2} P\{N(B_0) = 0 | \Theta = \theta\} f_\Theta(\theta) d\theta, \tag{9}$$

where $N(B_0)$ is the number of nodes located in B_0 and λ is the node density. As the nodes are randomly distributed, $N(B_i)$ can be modeled as a 2D Poisson point process [14]:

$$P\{N(B_i) = i | \Theta = \theta\} = \frac{(\lambda B_i)^i}{i!} e^{-\lambda B_i}. \tag{10}$$

And B_0 can be calculated as

$$B_0 = R^2 arccos\frac{lcos\theta}{R} - Rlcos\theta sin\left(arccos\frac{lcos\theta}{R}\right). \tag{11}$$

Assume Θ is uniformly distributed in $(-\pi/2, \pi/2)$,

$$F_{L_0}(l) = \int_{-\pi/2}^{\pi/2} e^{-\lambda B_0} \cdot \frac{1}{\pi} d\theta. \tag{12}$$

By differentiation, we can finally get $f_{L_0}(l)$ as follows:

$$f_{L_0}(l) = \frac{d}{dl} F_{L_0}(l) = -\frac{\lambda}{\pi} \int_{-\pi/2}^{\pi/2} e^{-\lambda B_0} \frac{\partial B_0}{\partial l} d\theta. \tag{13}$$

After the formulation of p_i, we can give the probability (for one hop) that when a packet is being sent from S, at least one of the $N+1$ potential forwarders[4] succeeds in receiving the packet (i.e., it is still within the transmission range of S) as follows:

$$P_{succ}^1 = \begin{cases} P\{N_F \geq 1\}(1 - p_0), & N = 0; \\ P\{N_F \geq 1\} \cdot P_{1hop}^N, & N \geq 1; \end{cases} \tag{14}$$

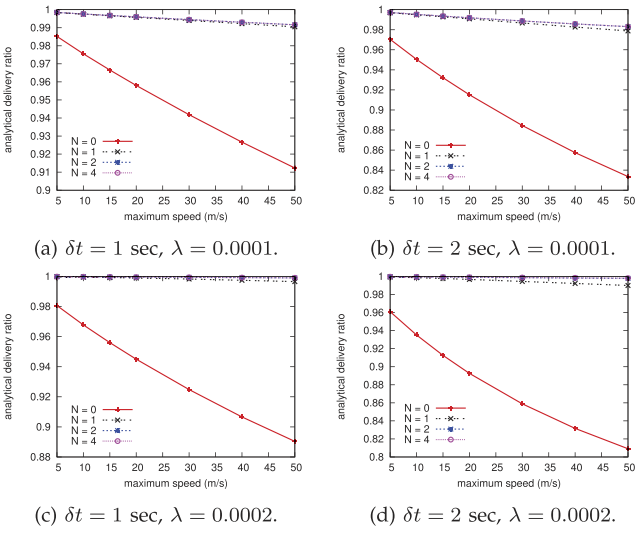
where N_F is the number of nodes that make positive progress toward the destination. $N = 0$ corresponds to the traditional ad hoc routing protocols where no forwarding candidates participate. P_{1hop}^N represents the probability that a packet can be successfully delivered when at most N ($N \geq 1$) forwarding candidates are involved and it is expressed as follows:

$$\begin{aligned} P_{1hop}^N = &\sum_{i=0}^{N-1} P\{N_C = i\} \left(1 - \prod_{j=0}^{i} p_j\right) \\ &+ P\{N_C \geq N\} \left(1 - \prod_{j=0}^{N} p_j\right), \end{aligned} \tag{15}$$

where N_C is the number of potential candidates located in the forwarding area (defined in Section 2.2). Similar to $N(B_0)$, both N_F and N_C can be modeled as a 2D Poisson point process. For simplicity, assume the area of the region within which nodes make progress toward the destination and the size of the forwarding area are $\pi R^2/2$ and $\pi(R/2)^2/2$, respectively, we can get $P\{N_F \geq 1\}$ as

$$P\{N_F \geq 1\} = 1 - P\{N_F = 0\} = 1 - e^{-\lambda \pi R^2/2}. \tag{16}$$

4. We manually set a maximum number N of forwarding candidates as implied in Algorithm 1. Therefore, the total number of potential forwarders is $N+1$, including the next hop. The effect of the value of N will be evaluated in the following parts.



(a) $\delta t = 1$ sec, $\lambda = 0.0001$. (b) $\delta t = 2$ sec, $\lambda = 0.0001$.

(c) $\delta t = 1$ sec, $\lambda = 0.0002$. (d) $\delta t = 2$ sec, $\lambda = 0.0002$.

Fig. 8. Analytical delivery ratio versus node mobility.

And $P\{N_C \geq N\}$ can be expressed as

$$P\{N_C \geq N\} = 1 - \sum_{i=0}^{N-1} \frac{(\lambda \pi R^2/8)^i}{i!} e^{-\lambda \pi R^2/8}. \tag{17}$$

The first term of (15) corresponds to the case in which the number of actual forwarding candidates is less than N, while the second term corresponds to the case in which the number of potential forwarding candidates is larger than N and we only select the best N nodes out of them.

For simplicity,[5] assume the per-hop packet forwarding is independent of each other, then the probability (denoted as analytical delivery ratio) that a packet is successfully delivered from the source to the destination under ideal wireless channel without route recovery scheme can be expressed as follows[6]:

$$P_{succ}^{N_h} = \left(P_{succ}^1\right)^{N_h}, \tag{18}$$

where N_h is the path length (number of hops).

According to the analytical formulas, we can see that node mobility (taking v_{max} as the metric), neighbor update interval δt, node density λ, and the maximum number of forwarding candidates N are four main factors affecting the value of $P_{succ}^{N_h}$. In the following part, we do some numerical evaluations to show the effect of these variables. Assume the path length N_h is 5 hops and the minimum absolute node speed v_{min} is 1 m/s, we get the analytical delivery ratio as a function of v_m with different values of N, δt and λ as shown in Fig. 8.

From Fig. 8, we can see that the involvement of forwarding candidates significantly improves the end-to-end delivery quality. The effect is most significant with the participation of the first candidate (i.e., $N = 1$). When $N \geq 2$,

5. Since the nodes are randomly and independently distributed in the whole area, without considering the effect of wireless interference (e.g., in light traffic case, as assumed previously), we can roughly accept the independence between hops to simplify our analysis.

6. Actually, the involvement of forwarding candidates will probably introduce a slightly longer end-to-end path length. Here, we ignore such factor for simplicity. The more accurate value of $P_{1hop}^{N_h}$ should thus be lower than that calculated according to (18).

the improvement seems to be marginal. One reason is that in most cases, two backup nodes are enough. On the other hand, the available forwarding candidates are restricted due to the selection of forwarding area. The delivery ratio $P_{succ}^{N_h}$ decreases significantly with the increased node mobility, while the failure ratio (i.e., $1 - P_{succ}^{N_h}$) almost doubles as we set δt from 1 to 2 s, showing the great impact of these two factors as we have argued above. Meanwhile, we find that the node density plays an interesting role in affecting the analytical delivery ratio. As the network becomes denser (e.g., Fig. 8a $\rightarrow$ 8c, 8b $\rightarrow$ 8d), $P_{succ}^{N_h}$ for $N = 0$ decreases due to the strategy used in greedy forwarding: with increased node density, the next hop selected is nearer to the boundary of the transmission range on average and thus it is easier to move out. However, the value of $P_{succ}^{N_h}$ for $N \geq 1$ increases on the contrary, which can be explained as follows: when the node density is high, the probability that there are at least N nodes located in the forwarding area (i.e., $P\{N_C \geq N\}$) becomes higher, and thus the advantage brought by in-the-air backup can be fully utilized. Such benefit overcomes the negative effect caused by the increased density and makes the delivery ratio larger than its counterpart in sparser networks.

4.2 Memory Consumption and Duplicate Relaying

One main concern of POR is its overhead due to opportunistic forwarding, as several copies of a packet need to be cached in the forwarding candidates, leading to more memory consumption, and duplicate relaying would possibly happen if the suppression scheme fails due to node mobility.[7] However, it will be presented later that this is not a serious problem.

We first look into the issue of memory consumption. If a packet is received by a forwarding candidate $C_{i,i\in[1,N]}$, it will be cached for a period of $i\Delta T$ at most according to the forwarding scheme described in Section 2.1, where ΔT is the time slot. Therefore, we can get the following upper bound for the length (number of packets cached) of the packet list Q_i at C_i for each flow:

$$Q_i \leq r_s \cdot i\Delta T, \tag{19}$$

where r_s is the packet sending rate at the source of the data flow. Suppose $r_s = 100$ packets/s (which is relatively heavy traffic); since we have set $\Delta T = 0.01$ s, Q_i would not exceed i, indicating that the opportunistic forwarding scheme used in our protocol will not consume much memory resource.

Then, we look into the overhead due to duplicate relaying. According to the analytical result presented in Section 4.1, $N = 2$ is enough to provide near optimal performance gain. Therefore in the following part, we consider the cases where at most two forwarding candidates are involved to keep our analysis concise. Different situations corresponding to different N will be discussed separately. Note that in this section, we assume the existence of enough nodes in the forwarding area (e.g., when considering $N = 2$, there are at least two nodes located in the forwarding area besides the next hop), so as to concentrate on the impact of the number of forwarding candidates and ignore the influence of node density. On the other hand, symmetric wireless link has been assumed by default, e.g., if node A can hear from node B, node B can also hear from node A. Since the communication time as well as the time slot is much shorter than the time scale of node mobility, if $C_{i,i\in[0,N]}$ can receive a packet from S, S can also be acknowledged (by C_0) or suppressed (by $C_{i,i\in[1,N]}$). The Expected Forwarding Times (EFT) are defined as the average number of times a packet is being forwarded and is used as the metric to evaluate duplicate forwarding.

4.2.1 $N = 0$ (No Forwarding Candidate Is Involved)

In this category, there are two possible cases: 1) the packet sent from S is successfully received by C_0, so it is forwarded only once; and 2) C_0 fails to receive the packet (i.e., it has moved out), then S reselects another next hop for this packet, and thus the packet is forwarded twice at this hop. Therefore, the expected forwarding times for $N = 0$ can be formulated as

$$EFT_0 = 1 \cdot \bar{p_0} + 2 \cdot p_0 = 1 + p_0, \tag{20}$$

where p_0 comes from (5) and $\bar{p_0} = 1 - p_0$ represents the probability that C_0 succeeds in receiving the packet.[8]

4.2.2 $N = 1$ (One Forwarding Candidate Is Involved)

In this category, the source of duplication is not only S's rerouting, but also C_1's duplicate relaying due to its moving out (i.e., C_1 is no longer within C_0's transmission range but is still within S's transmission range). The probability that the packet is forwarded once is

$$P_{1|N=1} = \bar{p_0}p_1 + p_0\bar{p_1} + \bar{p_0}\bar{p_1}P\{dist(C_0, C_1) \leq R|t\}, \tag{21}$$

where $P\{dist(C_0, C_1) \leq R|t\}$ denotes the probability that when a packet is sent from S at time t, the distance between C_0 and C_1 is smaller than R. The first two terms of (21) represent the cases that only C_0 or C_1 has received the packet, while the third term corresponds to the case that both C_0 and C_1 have received the packet but C_1 is suppressed by C_0 (i.e., $dist(C_0, C_1) \leq R$ at t) and thus no duplicate relaying is introduced. Since in this category a packet can only be forwarded once or twice, the probability that it is forwarded twice is actually complementary of $P_{1|N=1}$:

$$P_{2|N=1} = p_0p_1 + \bar{p_0}\bar{p_1}P\{dist(C_0, C_1) > R|t\}, \tag{22}$$

where the first term represents the case that neither C_0 nor C_1 has received the packet and S has to reroute the packet, while the second term corresponds to the case that C_1 is not suppressed by C_0 (as shown in Fig. 9a) and duplicate relaying is introduced.

Recall the definition of the forwarding area. When choosing a node as forwarding candidate, the distance between this node and C_0 should not exceed $R/2$ (as shown in Fig. 9a). So, we argue that $P\{dist(C_0, C_1) > R|t\}$ is close to zero and can be ignored in our formulation. Substitute $P\{dist(C_0, C_1) > R|t\} = 0$ into (21) and (22), we get the following approximate equations:

$$P_{1|N=1} \approx \bar{p_0}p_1 + p_0\bar{p_1} + \bar{p_0}\bar{p_1} = 1 - p_0p_1, \tag{23}$$

7. Actually, the suppression failure can also be caused by wireless fading or interference. In our analysis, we only consider the effect of node mobility as stated at the very beginning of this section.

8. Similar meaning for $\bar{p_1}$ and $\bar{p_2}$ in (21) to (27).

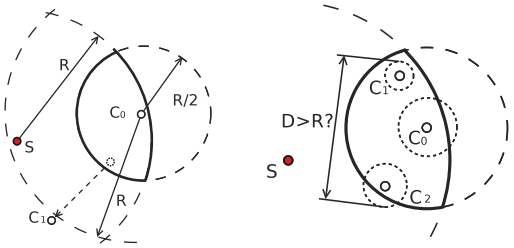


Fig. 9. (a) The case that C_1 is out of C_0's transmission range is rare. (b) C_1 and C_2 may be out of each others' transmission range when $N = 2$.

$$P_{2|N=1} \approx p_0 p_1. \tag{24}$$

The expected forwarding times for $N = 1$ can thus be expressed as follows:

$$EFT_1 = 1 \cdot P_{1|N=1} + 2 \cdot P_{2|N=1} \approx 1 + p_0 p_1. \tag{25}$$

Compare EFT_1 with EFT_0, we can see that the involvement of C_1 reduces the duplication instead of increasing it. Owing to the candidate selection policy, C_1 will be suppressed by C_0 with high probability and only marginal duplicate relaying will be introduced, while the backup bought about by C_1 will significantly decrease the duplication due to S's rerouting.

4.2.3 $N = 2$ (Two Forwarding Candidates Are Involved)

When two forwarding candidates are involved, we have to take duplicate relaying into more consideration. Though C_1 and C_2 will be suppressed by C_0 with high probability, in the case that C_0 moves out and C_1 forwards the packet instead, C_2 may not be successfully suppressed (as illustrated in Fig. 9b) since the initialized distance between C_1 and C_2 can be as far as R and they are much more likely to get separated (i.e., being outside each other's transmission range). Following the assumption that C_1 and C_2 will be suppressed by C_0 with probability 1 if either of them has overheard the packet, then the probability that a packet is forwarded once can be formulated as

$$\begin{aligned} P_{1|N=2} \approx\ & \bar{p_0} + p_0\bar{p_1}p_2 + p_0 p_1 \bar{p_2} \\ & + p_0\bar{p_1}\bar{p_2} P\{dist(C_1, C_2) \le R|t\}, \end{aligned} \tag{26}$$

where the first term represents the case that C_0 succeeds in receiving the packet and therefore we do not need to consider the overhearing of C_1 nor C_2 since no duplicate relaying will be introduced by assumption. The second and the third terms correspond to the cases that C_0 fails to receive the packet but only one of the forwarding candidates (C_1 or C_2) overhears it. The last term represents the case that C_0 fails to receive the packet while both C_1 and C_2 succeed in receiving it, but C_2 is suppressed by C_1, without introducing duplicate relaying. The probability that a packet is forwarded twice is complementary of $P_{1|N=2}$:

$$P_{2|N=2} \approx p_0 p_1 p_2 + p_0\bar{p_1}\bar{p_2} P\{dist(C_1, C_2) > R|t\}, \tag{27}$$

where the first term represents the case that none of the three nodes receives the packet (i.e., they all move out of the transmission range of S) and S has to reroute the packet, while the second term corresponds to the case that C_2 fails to be suppressed by C_1 when C_0 does not receive the packet, resulting in duplicate relaying. The expected forwarding times for $N = 2$ are given by

$$EFT_2 = 1 \cdot P_{1|N=2} + 2 \cdot P_{2|N=2}. \tag{28}$$

Let $p_0 = p_1 = p_2 = p$ and $P\{dist(C_1, C_2) > R|t\} = \alpha$, in which p and α reflect the degree of node mobility, we get the curves of EFT_0, EFT_1, and $EFT_2(\alpha)$ with varying p as shown in Fig. 10. It can be seen from the figure that when N is increased from 1 to 2, not much duplicate relaying is introduced (since 0.1 is already a large enough value for α). In addition, it can be further verified easily that EFT_2 would not be larger than EFT_0 under the assumption that $p_0 = p_1 = p_2 = p$.

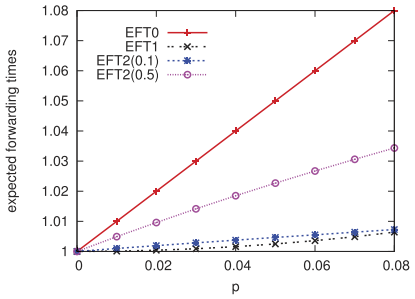


Fig. 10. EFT with different N.

5 Performance Evaluation

To evaluate the performance of POR, we simulate the algorithm in a variety of mobile network topologies in NS-2 [19] and compare it with AOMDV [20] (a famous multipath routing protocol) and GPSR [5] (a representative geographic routing protocol). The common parameters utilized in the simulations are listed in Table 2.

The improved random way point [21] without pausing is used to model nodes' mobility. The minimum node speed is set to 1 m/s and we vary the maximum speed to change the mobility degree of the network. The following metrics are used for performance comparison:

- *Packet delivery ratio*. The ratio of the number of data packets received at the destination(s) to the number of data packets sent by the source(s).
- *End-to-end delay*. The average and the median end-to-end delay are evaluated, together with the cumulative distribution function of the delay.
- *Path length*. The average end-to-end path length (number of hops) for successful packet delivery.

TABLE 2
Simulation Parameters

Parameter	Value
MAC Protocol	IEEE 802.11
Propagation Model	Two-ray Ground
Transmission Range	250 m
Mobility Model	Random Way Point (RWP)
Traffic Type	Constant Bit Rate (CBR)
Packet Size	256 Bytes
Number of Nodes	80
Simulation Time	900 sec

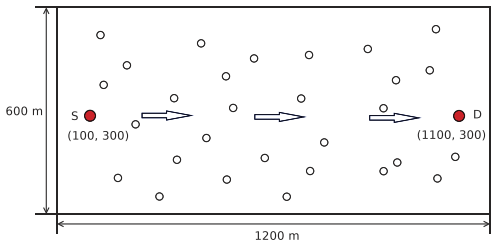


Fig. 11. Network topology: uniformly distributed.

- *Packet forwarding times per hop (FTH)*. The average number of times a packet is being forwarded from the perspective of routing layer to deliver a data packet over each hop.
- *Packet forwarding times per packet (FTP)*. The average number of times a packet is being forwarded from the perspective of routing layer to deliver a data packet from the source to the destination.

Among the metrics, FTH and FTP are designed to evaluate the amount of duplicate forwarding. For unicast-style routing protocols, packet reroute caused by path break accounts for FTH being greater than 1. On the other hand, for those packets who fail to be delivered to the destination(s), the efforts that have already been made in forwarding the packets are still considered in the calculation of FTH, as FTH is calculated as follows:

$$FTH = \frac{N_s + N_f}{\sum_{i=1}^{N_r} N_{hi}}, \tag{29}$$

where N_s, N_f, and N_r are the number of packets sent at the source(s), forwarded at intermediate nodes, and received at the destination(s), respectively. N_{hi} is the number of hops for the ith packet that is successfully delivered. Unlike FTH, FTP averages the total number of times a packet is being forwarded on a per-packet basis:

$$FTP = \frac{N_s + N_f}{N_r}. \tag{30}$$

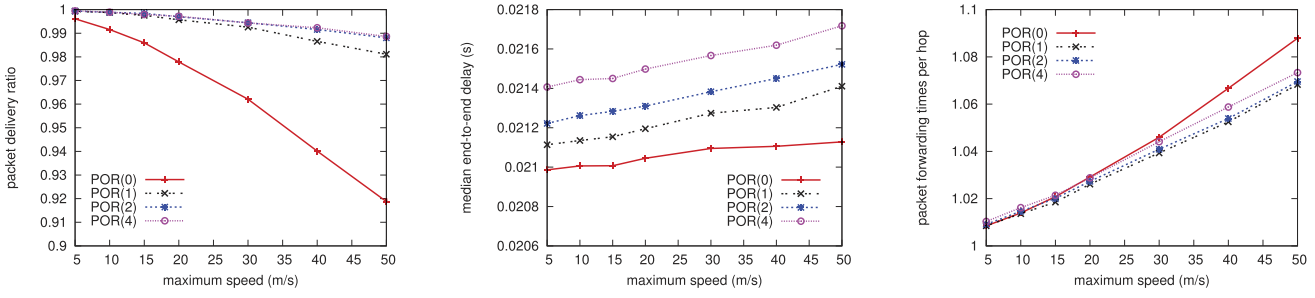
5.1 Forwarding Candidate Number Evaluation

We first evaluate the effect of the number of forwarding candidates (i.e., N) on POR's performance. Generally, larger value of N will result in higher robustness as more nodes serve as backups. However, it also means more memory resources need to be consumed and a higher percentage of duplicate relaying. In addition, the increase in the number of forwarding candidates will also enlarge the packet header, thus introducing more overhead. Therefore, a trade-off between the robustness and the required resource exists, in which the number of forwarding candidates plays an important role.

The network topology used in our simulation is illustrated in Fig. 11: 80 nodes are deployed in a rectangular area of size $1{,}200\,\mathrm{m} \times 800\,\mathrm{m}$. The source (S) and the destination (D) are fixed at the two ends to create a long enough end-to-end path length (≈ 5 hops), while the remaining 78 mobile nodes move according to the RWP model that we have described. A CBR flow is injected into the network at a rate of 10 packets per second (i.e., 20 Kbps), starting at 170 s and ending at 870 s. We vary the value of N from 0 to 4 and measure the packet delivery ratio, the median end-to-end delay, and the packet forwarding times per hop. The simulation results, averaged over 10 independent runs, are shown in Fig. 12.

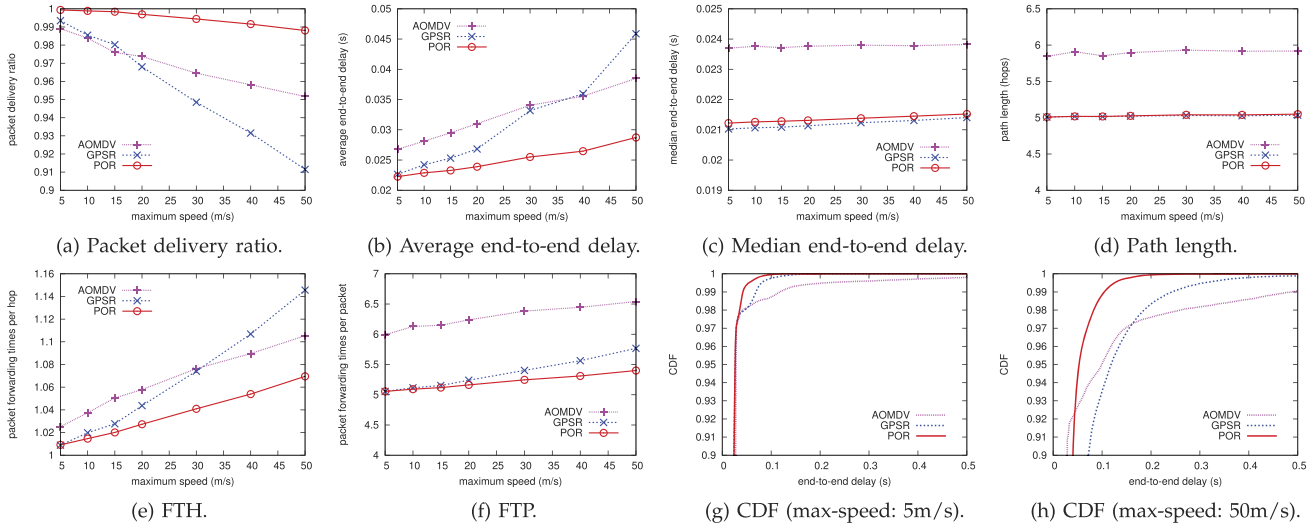
From Fig. 12a, we can see that though more forwarding candidates yield a higher packet delivery ratio, only the involvement of the first forwarding candidate achieves the most significant performance gain, while the improvement becomes less and less observable when N continues to increase, which is consistent with our theoretical analysis presented in Section 4.1. Note that in the operation of routing protocols when link break happens, some recovery scheme (e.g., packet rerouting) will be triggered to salvage the packet. Hence, the simulated delivery ratio tends to be higher than the analytical one, especially for the protocol without forwarding candidates (i.e., POR(0)). On the other hand, the measured result should be lower than the analytical one due to the impact of wireless interference on the contrary. These two factors, together with ignoring the change of the path length (as mentioned in footnote 6), contribute to the difference between the simulated delivery ratio and the analytical delivery ratio.

Fig. 12b shows that the median end-to-end delay grows more or less linearly with the number of forwarding candidates. The reason is due to the increased packet size as the IP addresses of the forwarding candidates are attached to the packet header. Pertaining to duplicate relaying (Fig. 12c), we can see that the involvement of forwarding candidates reduces the value of FTH instead of introducing more duplication, especially when the node mobility is high, as shown in Section 4.2. Note that the calculation of FTH also takes the wasted forwarding of lost packets into consideration. Thus, the improvement in packet delivery ratio also contributes to the reduction of FTH. When more candidates



(a) Packet delivery ratio. (b) Median end-to-end delay. (c) FTH.

Fig. 12. Forwarding candidate number evaluation.



(a) Packet delivery ratio. (b) Average end-to-end delay. (c) Median end-to-end delay. (d) Path length.

(e) FTH. (f) FTP. (g) CDF (max-speed: 5m/s). (h) CDF (max-speed: 50m/s).

Fig. 13. Simulation results: uniformly distributed, single flow.

participate in packet forwarding, FTH grows gradually as expected.

Based on the simulation results together with our analysis, we make the following conclusion: in most cases, two candidates are enough to provide high robustness while keeping the overhead relatively low. Thus in the following experiments, we all set the number of forwarding candidates as 2.

5.2 Effect of Mobility: Single-Flow Case

To evaluate the effect of mobility on the performance of routing protocols, we first consider the single-flow case in uniformly distributed network. The same simulation scenario as that of Section 5.1 (Fig. 11) is used to create a path length controllable dynamic topology. The results based on 10 independent runs are listed in Fig. 13.

Fig. 13a highlights the effectiveness of the backup in the air utilized by POR. Even when the maximum node speed increases to 50 m/s, POR still enables nearly 99 percent of the packets to reach the destination while the delivery ratios of AOMDV and GPSR both decrease significantly. It can be observed that when the maximum node speed is not larger than 15 m/s, GPSR still outperforms AOMDV but the reverse is true when node mobility keeps increasing. The reason is due to the location sensitivity of GPSR, which becomes more serious when the path length is large as the routing decision is made hop by hop, resulting in a multiplier effect. For AOMDV, though multiple paths are exploited for backup, it still fails to maintain high packet delivery ratio in the face of highly dynamic network topology.

From Figs. 13b and 13c, we can see that POR is not only effective but also efficient. It delivers as many as possible packets at extremely low delay. The near optimum path length (Fig. 13d) contributes to the efficiency. On the other hand, the mitigation of route recovery because of forwarding candidates' collaboration reduces the end-to-end delay significantly. Especially in highly dynamic network, frequent occurrence of link breaks due to node mobility would introduce substantial latency for the immediately affected few packets, which can be as long as several seconds. These packets will dominate the average delay. The rare occurrence, if any, of such highly delayed packets in POR underscores POR's prominent performance as shown in Fig. 13b where POR can efficiently maintain uninterrupted communication. With respect to the CDF of latency (Figs. 13g and 13h), the curve for POR converges to 100 percent very quickly while in the other two protocols, there is a certain number of packets experiencing extremely long delay.

As for the overhead, Figs. 13e and 13f show that POR's excellent performance is not at the cost of increased duplication. On the contrary, both the FTH and FTP of POR are the lowest, further verifying POR's outstanding efficiency. From the result, we can conclude that POR achieves the highest bandwidth utilization among the three routing protocols.

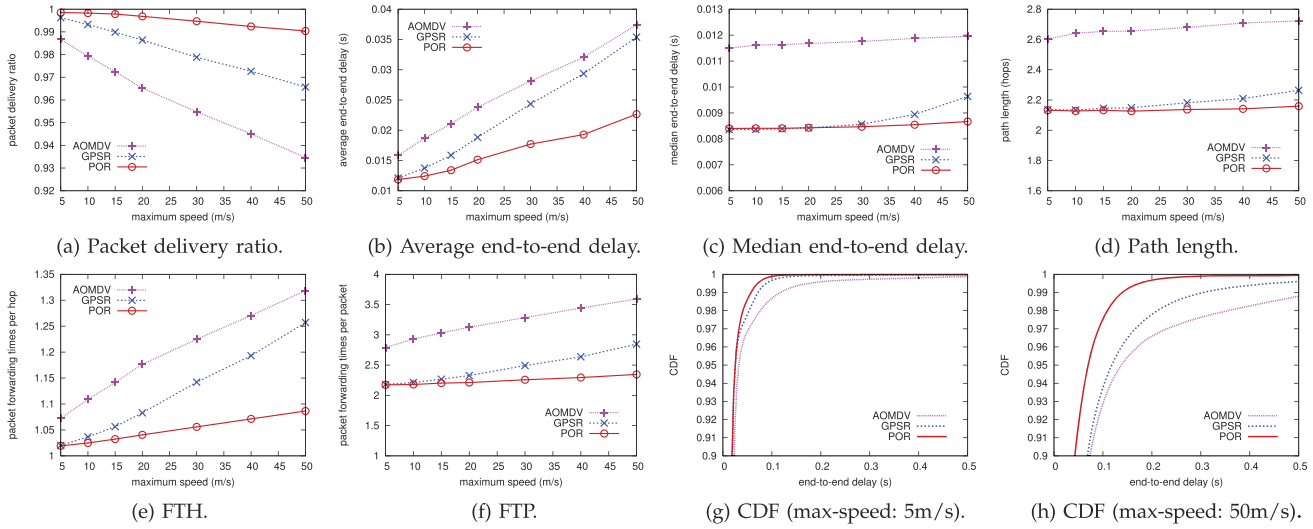
As a summary, POR outperforms AOMDV and GPSR in packet delivery ratio, end-to-end delay, as well as resource (bandwidth) efficiency. Benefiting from the per-hop routing decision and thus stateless property, GPSR seems to be better than AOMDV when the node mobility is not so high but fails to keep up the performance when node mobility exceeds a certain threshold.

5.3 Effect of Mobility: Multiflow Case

To fully evaluate the performance of POR in the face of node mobility, the multiflow case is also taken into consideration. In our simulation setup, 80 mobile nodes are located in a $1{,}200\ \mathrm{m} \times 600\ \mathrm{m}$ rectangular region, similar to that used in the previous experiments, but all the nodes move according to the RWP model. For traffic, 10 CBR flows are simulated and each flow sends at a rate of 4 Kbps. The results are illustrated in Fig. 14.

Comparing with Fig. 13, we find that the difference is not distinct. POR still enjoys a clear win over the other two. Owing to multiple flows, the curves of CDF are smoother. Interestingly, we find that GPSR outperforms AOMDV. The main reason is due to the shortened path length.[9] The

9. Note that in the multiflow case, the source and destination pairs are randomly chosen from the mobile nodes which are evenly distributed in the whole simulation area (actually, nodes tend to be concentrated in the central area when using RWP as the mobility model [17]). So on average, the path length is shorter than that in the single-flow case.



(a) Packet delivery ratio. (b) Average end-to-end delay. (c) Median end-to-end delay. (d) Path length.

(e) FTH. (f) FTP. (g) CDF (max-speed: 5m/s). (h) CDF (max-speed: 50m/s).

Fig. 14. Simulation results: uniformly distributed, multiflow.

multiplier effect caused by inaccurate location information is alleviated and the performance gain from hop-by-hop routing makes GPSR more tolerant to node mobility than AOMDV in the shorter path length scenario.

5.4 Effect of Communication Hole

To test the effectiveness of VDVH, we further evaluate the routing performance in mobile networks with a communication hole. We create a network topology as illustrated in Fig. 15. The source and destination nodes are fixed at the two ends of the rectangle while the remaining 78 nodes move in the annular region according to the RWP model. The central gray area is simulated as the communication hole with no mobile node distributed. The traffic setup is the same as that in Section 5.1. By changing the maximum node speed, we obtain the simulation results shown in Fig. 16.

From Fig. 16a, we can observe that in the face of communication hole, GPSR's void handling mechanism fails to work well. Even when the maximum node speed is 5 m/s, only 90 percent of the data packets get delivered which is relatively poor compared to the other protocols. As for POR, the improvement is not so significant since in the current implementation, VDVH is unable to deal with all cases of communication voids. However, when the node mobility is high (e.g., when the maximum node speed is larger than 25 m/s), POR still performs better.

With respect to the path length, the end-to-end hops of GPSR are the largest due to the usage of perimeter mode,

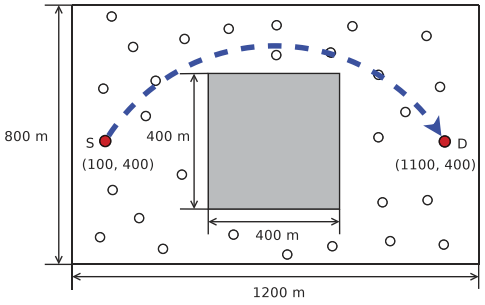


Fig. 15. Network topology: with communication hole.

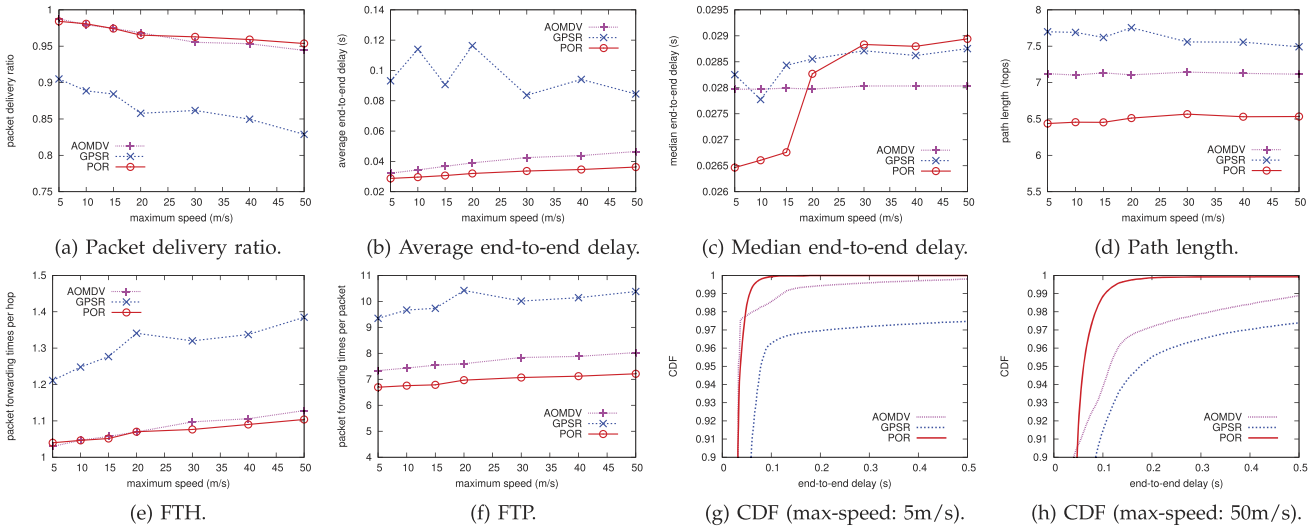
while POR still achieves the shortest path length (Fig. 16d), leading to the low latency (Figs. 16b, 16c, 16g, and 16h). As for the result of FTH and FTP, POR outperforms the other two as usual while GPSR performs worst indicating that the perimeter mode of GPSR is incapable of working well in mobile environment. Our new void handling scheme, though simple, is quite tolerant to node mobility and is endowed with the capability of finding the shortest path around the communication hole.

6 Related Work

To enhance a system's robustness, the most straightforward method is to provide some degree of redundancy. According to the degree of redundancy, existing robust routing protocols for MANETs can be classified into two categories. One uses the *end-to-end redundancy*, e.g., multipath routing, while the other leverages on the *hop-by-hop redundancy* which takes advantage of the broadcast nature of wireless medium and transmits the packets in an opportunistic or cooperative way. Our scheme falls into the second category.

Multipath routing, which is typically proposed to increase the reliability of data transmission [22] in wireless ad hoc networks, allows the establishment of multiple paths between the source and the destination. Existing multipath routing protocols are broadly classified into the following three types: 1) using alternate paths as backup (e.g., [20], [23], [24]); 2) packet replication along multiple paths (e.g., [13], [25]); and 3) split, multipath delivery, and reconstruction using some coding techniques (e.g., [26], [27]). However, as discussed in [28], it may be difficult to find suitable number of independent paths. More importantly, in the face of high node mobility, all paths may be broken with considerably high probability due to constantly changing topology, especially when the end-to-end path length is long, making multipath routing still incapable of providing satisfactory performance.

In recent years, wireless broadcast is widely exploited to improve the performance of wireless communication. The concept of opportunistic forwarding, which was used to



(a) Packet delivery ratio. (b) Average end-to-end delay. (c) Median end-to-end delay. (d) Path length.

(e) FTH. (f) FTP. (g) CDF (max-speed: 5m/s). (h) CDF (max-speed: 50m/s).

Fig. 16. Simulation results: with communication hole.

increase the network throughput ([6], [7]), also shows its great power in enhancing the reliability of data delivery.

In the context of infrastructure networks, by using opportunistic overhearing, the connectivity between the mobile node and base station (BS) can be significantly improved. In [29], an opportunistic retransmission protocol PRO is proposed to cope with the unreliable wireless channel. Implemented at the link layer, PRO leverages on the path loss information Receiver Signal Strength Indicator (RSSI) to select and prioritize relay nodes. By assigning the higher priority relay a smaller contention window size, the node that has higher packet delivery ratio to the destination will be preferred in relaying. With respect to the impact of mobility, Wu et al. [30] investigate the WiFi connectivity for moving vehicles, with focus on the cooperation among BSs. BSs that overhear a packet but not its acknowledgment probabilistically relay the packet to the intended next hop. With the help of auxiliary BSs, the new protocol performs much better than those schemes with only one BS participating in the communication even if advanced link prediction and handover methods are involved. However, due to the lack of strict coordination between BSs, false positives and false negatives exist.

While the aforementioned two schemes deal with the issues in WLANs, the authors of [31] concentrate on the robust routing in mobile wireless sensor networks. In the proposed RRP, traditional ad hoc routing mechanism is used to discover an intended path while the nodes nearby act as guard nodes. Leveraging on a modified 802.11 MAC, guard nodes relay the packet with prioritized backoff time when the intended node fails. If the failure time exceeds a certain threshold, the guard node who has recently accomplished the forwarding will become the new intended node. A potential problem is that such substitution scheme may lead to suboptimal paths. Unlike RRP, our protocol uses location information to guide the data flow and can always archive near optimal path. On the other hand, our scheme focuses on the route discovery from the perspective of network layer and no such complex MAC modification is necessary. Forwarding candidates are coordinated using the candidate list and no contention would happen. By limiting the forwarding area, duplication can also be well controlled.

7 Conclusions

In this paper, we address the problem of reliable data delivery in highly dynamic mobile ad hoc networks. Constantly changing network topology makes conventional ad hoc routing protocols incapable of providing satisfactory performance. In the face of frequent link break due to node mobility, substantial data packets would either get lost, or experience long latency before restoration of connectivity. Inspired by opportunistic routing, we propose a novel MANET routing protocol POR which takes advantage of the stateless property of geographic routing and broadcast nature of wireless medium. Besides selecting the next hop, several forwarding candidates are also explicitly specified in case of link break. Leveraging on such natural backup in the air, broken route can be recovered in a timely manner. The efficacy of the involvement of forwarding candidates against node mobility, as well as the overhead due to opportunistic forwarding is analyzed. Through simulation, we further confirm the effectiveness and efficiency of POR: high packet delivery ratio is achieved while the delay and duplication are the lowest.

On the other hand, inherited from geographic routing, the problem of communication void is also investigated. To work with the multicast forwarding style, a virtual destination-based void handling scheme is proposed. By temporarily adjusting the direction of data flow, the advantage of greedy forwarding as well as the robustness brought about by opportunistic routing can still be achieved when handling communication voids. Traditional void handling method performs poorly in mobile environments while VDVH works quite well.

Acknowledgments

This work was supported by the Agency for Science, Technology, and Research (A*STAR) of Singapore (Grant number: 0721010028 - M47020034).

REFERENCES

[1] J. Broch, D.A. Maltz, D.B. Johnson, Y.-C. Hu, and J. Jetcheva, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols," *Proc. ACM MobiCom,* pp. 85-97, 1998.

[2] M. Mauve, A. Widmer, and H. Hartenstein, "A Survey on Position-Based Routing in Mobile Ad Hoc Networks," *IEEE Network,* vol. 15, no. 6, pp. 30-39, Nov./Dec. 2001.

[3] D. Chen and P. Varshney, "A Survey of Void Handling Techniques for Geographic Routing in Wireless Networks," *IEEE Comm. Surveys and Tutorials,* vol. 9, no. 1, pp. 50-67, Jan.-Mar. 2007.

[4] D. Son, A. Helmy, and B. Krishnamachari, "The Effect of Mobility Induced Location Errors on Geographic Routing in Mobile Ad Hoc Sensor Networks: Analysis and Improvement Using Mobility Prediction," *IEEE Trans. Mobile Computing,* vol. 3, no. 3, pp. 233-245, July/Aug. 2004.

[5] B. Karp and H.T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," *Proc. ACM MobiCom,* pp. 243-254, 2000.

[6] S. Biswas and R. Morris, "EXOR: Opportunistic Multi-Hop Routing for Wireless Networks," *Proc. ACM SIGCOMM,* pp. 133-144, 2005.

[7] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading Structure for Randomness in Wireless Opportunistic Routing," *Proc. ACM SIGCOMM,* pp. 169-180, 2007.

[8] E. Rozner, J. Seshadri, Y. Mehta, and L. Qiu, "SOAR: Simple Opportunistic Adaptive Routing Protocol for Wireless Mesh Networks," *IEEE Trans. Mobile Computing,* vol. 8, no. 12, pp. 1622-1635, Dec. 2009.

[9] A. Balasubramanian, R. Mahajan, A. Venkataramani, B.N. Levine, and J. Zahorjan, "Interactive WiFi Connectivity for Moving Vehicles," *Proc. ACM SIGCOMM,* pp. 427-438, 2008.

[10] K. Zeng, Z. Yang, and W. Lou, "Location-Aided Opportunistic Forwarding in Multirate and Multihop Wireless Networks," *IEEE Trans. Vehicular Technology,* vol. 58, no. 6, pp. 3032-3040, July 2009.

[11] S. Das, H. Pucha, and Y. Hu, "Performance Comparison of Scalable Location Services for Geographic Ad Hoc Routing," *Proc. IEEE INFOCOM,* vol. 2, pp. 1228-1239, Mar. 2005.

[12] R. Flury and R. Wattenhofer, "MLS: An Efficient Location Service for Mobile Ad Hoc Networks," *Proc. ACM Int'l Symp. Mobile Ad Hoc Networking and Computing (MobiHoc),* pp. 226-237, 2006.

[13] E. Felemban, C.-G. Lee, E. Ekici, R. Boder, and S. Vural, "Probabilistic QoS Guarantee in Reliability and Timeliness Domains in Wireless Sensor Networks," *Proc. IEEE INFOCOM,* pp. 2646-2657, 2005.

[14] D. Chen, J. Deng, and P. Varshney, "Selection of a Forwarding Area for Contention-Based Geographic Forwarding in Wireless Multi-Hop Networks," *IEEE Trans. Vehicular Technology,* vol. 56, no. 5, pp. 3111-3122, Sept. 2007.

[15] N. Arad and Y. Shavitt, "Minimizing Recovery State in Geographic Ad Hoc Routing," *IEEE Trans. Mobile Computing,* vol. 8, no. 2, pp. 203-217, Feb. 2009.

[16] Y. Han, R. La, A. Makowski, and S. Lee, "Distribution of Path Durations in Mobile Ad-Hoc Networks - Palm's Theorem to the Rescue," *Computer Networks,* vol. 50, no. 12, pp. 1887-1900, 2006.

[17] W. Navidi and T. Camp, "Stationary Distributions for the Random Waypoint Mobility Model," *IEEE Trans. Mobile Computing,* vol. 3, no. 1, pp. 99-108, Jan./Feb. 2004.

[18] R. Groenevelt, "Stochastic Models for Mobile Ad Hoc Networks," PhD dissertation, Universite de Nice, Sophia Antipolis, France, 2005.

[19] The Network Simulator ns-2, http://www.isi.edu/nsnam/ns, 2011.

[20] M. Marina and S. Das, "On-Demand Multipath Distance Vector Routing in Ad Hoc Networks," *Proc. Ninth Int'l Conf. Network Protocols (ICNP '01),* pp. 14-23, Nov. 2001.

[21] J. Yoon, M. Liu, and B. Noble, "Random Waypoint Considered Harmful," *Proc. IEEE INFOCOM,* pp. 1312-1321, 2003.

[22] S. Mueller, R. Tsang, and D. Ghosal, "Multipath Routing in Mobile Ad Hoc Networks: Issues and Challenges," *Performance Tools and Applications to Networked Systems,* pp. 209-234, Springer, 2004.

[23] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly Resilient, Energy-Efficient Multipath Routing in Wireless Sensor Networks," *ACM SIGMOBILE Mobile Computing and Comm. Rev.,* vol. 5, no. 4, pp. 11-25, 2001.

[24] A. Valera, W. Seah, and S. Rao, "Improving Protocol Robustness in Ad Hoc Networks through Cooperative Packet Caching and Shortest Multipath Routing," *IEEE Trans. Mobile Computing,* vol. 4, no. 5, pp. 443-457, Sept./Oct. 2005.

[25] B. Deb, S. Bhatnagar, and B. Nath, "ReInForM: Reliable Information Forwarding Using Multiple Paths in Sensor Networks," *Proc. Ann. IEEE Int'l Conf. Local Computer Networks (LCN '03),* pp. 406-415, 2003.

[26] A. Tsirigos and Z. Haas, "Analysis of Multipath Routing-Part I: The Effect on the Packet Delivery Ratio," *IEEE Trans. Wireless Comm.,* vol. 3, no. 1, pp. 138-146, Jan. 2004.

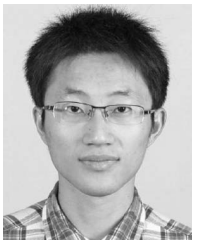
[27] A. Tsirigos and Z. Haas, "Analysis of Multipath Routing, Part 2: Mitigation of the Effects of Frequently Changing Network Topologies," *IEEE Trans. Wireless Comm.,* vol. 3, no. 2, pp. 500-511, Mar. 2004.

[28] Z. Ye, S. Krishnamurthy, and S. Tripathi, "A Framework for Reliable Routing in Mobile Ad Hoc Networks," *Proc. IEEE INFOCOM,* pp. 270-280, 2003.

[29] M.-H. Lu, P. Steenkiste, and T. Chen, "Design, Implementation and Evaluation of an Efficient Opportunistic Retransmission Protocol," *Proc. ACM MobiCom,* pp. 73-84, 2009.

[30] F. Wu, T. Chen, S. Zhong, L.E. Li, and Y.R. Yang, "Incentive-Compatible Opportunistic Routing for Wireless Networks," *Proc. ACM MobiCom,* pp. 303-314, 2008.

[31] X. Huang, H. Zhai, and Y. Fang, "Robust Cooperative Routing Protocol in Mobile Wireless Sensor Networks," *IEEE Trans. Wireless Comm.,* vol. 7, no. 12, pp. 5278-5285, Dec. 2008.



Shengbo Yang received the BE degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2008. He is currently working toward the PhD degree at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include routing in challenged mobile ad hoc networks, delay tolerant networks, and vehicular networks.

Chai Kiat Yeo received the BE (Hons) and MSc degrees in 1987 and 1991, respectively, both in electrical engineering, from the National University of Singapore, and the PhD degree from the School of Electrical and Electronics Engineering, Nanyang Technological University (NTU), Singapore, in 2007. She was a principal engineer with Singapore Technologies Electronics and Engineering Limited prior to joining NTU in 1993. She has been the deputy director of the Centre for Multimedia and Network Technology (CeMNet) at Nanyang Technological University. She is currently an associate chair (Academic) with the School of Computer Engineering, NTU. Her research interests include ad hoc and mobile networks, overlay networks, and speech processing and enhancement.

Bu Sung Lee received the BSc (Hons) and PhD degrees from the Electrical and Electronics Department, Loughborough University of Technology, United Kingdom, in 1982 and 1987, respectively. He is currently an associate professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He was elected the inaugural president of Singapore Research and Education Networks (SingAREN) and has been an active member of several national standards organizations, including the National Grid Pilot Project and international organizations such as Asia Pacific Advanced Networks (APAN). His research interests include computer networks protocols, distributed computing, network management, and grid computing.