Joint Design of Routing and Medium Access Control for Hybrid Mobile Ad Hoc Networks

Xiaojiang (James) Du · Dapeng Wu

Abstract Efficient routing and medium access control (MAC) are very important for Mobile Ad hoc Networks (MANETs). Most existing routing and MAC protocols consider homogeneous ad hoc networks, in which all nodes are modeled as the same, i.e., they have the same communication capabilities and characteristics. Although a homogeneous network model is simple and easy to analyze, it misses important characteristics of many realistic MANETs such as military battlefield ad hoc networks. In addition, a homogeneous ad hoc network suffers from poor performance and scalability. In many ad hoc networks, multiple types of nodes do co-exist; and some nodes have larger transmission power, higher transmission data rate, and better processing capability, are more reliable and robust than other nodes. Hence, a hybrid network model is more realistic and provides many advantages for designing better routing and MAC protocols. In this paper, we present a new routing protocol called Hybrid Routing, which is specifically designed for hybrid MANETs. In addition, a novel MAC protocol is jointly designed for hybrid MANETs. Extensive simulations demonstrate that the proposed routing MAC protocols have very good performance.

X. Du
Department of Computer Science, North Dakota State University,
Fargo, ND 58105, USA
e-mail: Xiaojiang.Du@ndsu.edu

D. Wu (✉)
Department of Electrical and Computer Engineering,
University of Florida, Gainesville,
Florida, FL 32611, USA
e-mail: wu@ece.ufl.edu

1 Introduction

Mobile Ad hoc Networks are dynamic multi-hop wireless networks that are established by a group of mobile nodes on shared wireless channel. It is characterized by no fixed infrastructure, dynamic topologies, bandwidth-constrained, variable capacity links, energy constrained operation and limited physical security.

Routing in MANETs has been extensively studied over the past several years [2, 4, 6, 7, 9, 11], and many ad hoc routing protocols have been proposed. Most existing routing protocols assume homogeneous MANETs. That is, all nodes have the same communication capabilities and characteristics—the same transmission power (range), transmission data rate, reliability and security. However, a homogeneous ad hoc network suffers from poor performance and scalability. Recent research has demonstrated its performance bottleneck both theoretically [1] and through simulation experiments and testbed measurement [2]. Gupta and Kumar [1] showed that per node capacity of a homogeneous wireless network is only $\Theta\left(\frac{W}{\sqrt{n \log n}}\right)$ (where W is the node transmission capacity and n is the number of nodes). Furthermore, in many realistic ad hoc networks, nodes are hybrid. For example, in a battlefield network, there are solders carrying portable wireless devices, there are vehicles and tanks carrying more powerful and reliable communication devices, and there may be aircrafts and satellites flying above and cover the whole battlefield. They have different communication characteristics in terms of transmission power, data rate, processing capability, reliability,

etc. So it would be more realistic to model these network elements as different types of nodes. Also there are many advantages that can be utilized to design better routing protocol when nodes are modeled as different types.

In this paper, we present a new routing protocol—Hybrid Routing (HR) protocol for hybrid MANETs. HR achieves good performance by exploiting node heterogeneity in many MANETs. Also, node location information is used to reduce routing overhead. The rest of the paper is organized as below. Section 2 discusses the related work. Section 3 presents our HR protocol. In Section 4, we compare the HR protocol with Ad hoc On Demand Distance Vector (AODV) [7] routing through extensive simulations. Section 5 concludes the paper.

2 Related work

Since many routing protocols have been proposed for MANETs, first we want to distinguish our work from the existing works. Our HR protocol takes advantage of different communication capabilities of heterogeneous nodes in many ad hoc networks; specifically, some physically more powerful nodes are chosen as backbone nodes for routing. Although the idea of using backbone nodes in routing is not new, our HR protocol is unique in backbone node selection and routing scheme. Next, we compare the HR with some related works.

The CEDAR [13] algorithm establishes and maintains a routing infrastructure called *core* in ad hoc networks. And routing is based on the *core*. There are several differences between CEDAR and our HR protocol protocol. We list some of the differences in the following: CEDAR considers homogeneous nodes, while HR considers hybrid nodes. The hybrid node model is more realistic and provides efficient routing. In CEDAR, some complex algorithm is used to generate and maintain the core nodes, and the algorithm introduces large overhead, since every node needs to broadcast messages to its neighbors periodically. While in HR, the election of backbone nodes is very simple, the first backbone-capable (more powerful) node that sends out claim message becomes the backbone node. In addition, CEDAR needs to broadcast route probe packet to discover the location of a destination node. While in HR, Global Positioning System (GPS) is used to provide node location information, and an efficient algorithm is used to disseminate node location information.

In [2], Xu et al. proposed MBN routing protocol with backbone nodes. The major differences between our HR protocol and MBN are the way to deploy backbone nodes and the routing algorithm for backbone nodes. In MBN, multi-hop clustering scheme is used to form clusters in the network, and the cluster heads become the backbone nodes.

However, the multi-hop clustering algorithm is complex and often causes concurrent cluster head claims. In HR protocol, the backbone node deployment is based on node location information. The entire routing area is divided into several small equal-size squares—cells. And one backbone node is elected in each cell. A simple algorithm is used for backbone node election. In MBN, routing among backbone nodes is based on another routing algorithm—LANMAR [17], which is not trivial. Furthermore, LANMAR uses logical group concept to aid routing. However, the logical group is not applicable to all MANETs. In HR protocol, routing among backbone nodes is very simple. Based on node location information and cell, it is very easy to determine a short path from a source to a destination, and no algorithm is needed. The detail is given in Section 3.

Ye et al. considered non-homogeneous MANETs in [6], and they proposed to place reliable nodes to assist routing. In HR protocol, we do not consider placing additional reliable nodes in the network. Instead, we utilize existing nodes in hybrid MANETs, and the powerful nodes in HR protocol have limited transmission range. There are also some papers that deal with node heterogeneity [14–16]. However, they mainly discuss how to solve the unidirectional link problem in ad hoc networks. In HR protocol, we consider how to take advantage of the different communication capabilities of hybrid nodes and provide better routing strategy. The unidirectional link problem also exists in HR protocol, i.e., the connection from source or destination to nearby backbone node. Since usually the source (or destination) is in the same cell as the nearest backbone node, it is only a few hops to the backbone node. We solve this unidirectional link problem in HR protocol as following: When there is a packet needs to be sent, source (or destination) node floods a route request packet within the cell to find out the path to the nearest backbone node.

Research has shown that geographical location information can improve routing performance in ad hoc networks. Routing with assistance from geographic location information requires each node to be equipped with GPS device. This requirement is quite realistic today since such devices are inexpensive and can provide reasonable precision. Several routing algorithms based on location information have been proposed. The well-known location based routing algorithms are Location-Aided Routing (LAR) protocol [3], Distance Routing Effect Algorithm for Mobility (DREAM) [10], Greedy Perimeter Stateless Routing (GPSR) [5], etc. HR protocol utilizes node location information to simplify routing strategy. The entire routing area is divided into several cells. The cell or grid concept is also used in GRID [9]. There are several differences between HR and GRID. The major difference is that HR considers hybrid ad hoc networks, while GRID considers homogeneous ad hoc networks.

3 Multi-class routing

Many ad hoc networks are hybrid MANETs, where physically different kinds of nodes present. Thus it would be more realistic to model nodes in such networks as different types of nodes. For simplicity, we consider there are only two types of nodes in the network. One type of node is high-end node and has larger transmission range (power), higher data rate, better processing capability, and is more reliable and robust than the other type. We refer to the more powerful nodes as Backbone-Capable nodes, in short as BC-nodes. In HR protocol, BC-nodes can be elected and serve as the Backbone nodes (B-nodes). Other nodes are referred to as general nodes. It is easy to extend HR protocol to consider more than two types of nodes. In this paper, we only consider the two-type case.

The main idea of HR protocol is to let most routing activities rely on B-nodes. There are several advantages of routing based on B-nodes: (1) It provides better reliability and fault tolerance, since B-nodes are more reliable than general nodes. (2) B-nodes have higher data rate than general nodes. Routing packets via B-nodes is more efficient than using general nodes. (3) B-nodes have larger transmission range, which reduces the number of hops in routing, thus reduces the routing overhead and latency.

Usually the transmission range of B-nodes—R is much larger than that of general nodes—r. The routing area is divided into several small, equal-sized squares—referred to as cells, as shown in Fig. 2. If the side length of a cell is set as $a = R/(2\sqrt{2})$, as shown in Fig. 1, where it is the worst case (longest distance) between two neighbor B-nodes, then a B-node can always communicate directly with B-nodes in all nearby cells, including the diagonal one. Since most of the time, two nearby B-nodes are not in the two opposite corners, we can use a larger cell size with $a > R/(2\sqrt{2})$, and still ensure the connection of nearby B-nodes for most of the time. More detailed discussion of cell size can be found in [8]. One and only one B-node is elected and maintained in each cell if there are several BC-nodes available in the cell. In HR protocol, we assume the routing area is fixed, i.e., nodes move around in a fixed territory. This is true for many MANETs (e.g., military battlefield networks, disaster relieve networks, conferences, conven-

Fig. 1 The relationship between a and R

tion centers, and so on). They all have a fixed routing territory. Since the routing area is fixed, the position of each cell is also fixed. Given the location (coordinates) of a node, there is a predefined mapping between the node location and the cell it is in.

The rest of this section is organized as below. Section 3.1 presents the Hybrid Routing protocol. In Section 3.2, we discuss the B-node election algorithm. In Section 3.3, we give an example to illustrate how HR protocol works. In Section 3.4, we present a novel MAC protocol to leverage the different transmission capabilities of B-nodes and G-nodes which further improves the MAC-layer performance.

3.1 The hybrid routing protocol

The HR protocol is presented in the following.

1. There is a unique id for each cell. In Fig. 2, the number is the id for each cell. One (and only one) B-node is elected and maintained in each cell, and each B-node has a second address, which is the same as the id of the cell in which it stays. Thus a B-node can send packet to a B-node in a nearby cell by using the second address, even though the identity of that B-node may change.
2. When a B-node moves out of a cell, it initiates a B-node election process in the cell. And a BC-node will be elected as the new B-node. The B-node election algorithm is described in the sequel.
3. Routing among B-nodes. First we discuss how B-nodes find routes to other B-nodes. B-nodes use their second addresses to communicate with each other. Assume B-node B_S (in cell C_S) wants to send a packet to B-node B_d (in cell C_d). Although nodes move around, the cells are fixed. B_S knows the cell in which B_d stays based on B_d's second address, which is the same as the cell id. A straight line L is drawn between the centers of cell C_S and C_d. An example is given in Fig. 2, assume B-node 7 wants to send a packet to B-node 3. The center (red) line is line L. Two border lines (blue lines) which parallel to line L with distance of W from L are drawn from C_S to C_d. All the cells that are within the two border lines are defined as *routing cells*. The value of W depends on the density of BC-nodes in the network. If there are enough BC-nodes in the network, i.e., with high probability there is at least one BC-node in each cell, then W can be set as zero, i.e., *routing cells* are only the cells that intercept with red line L—cell 7, 5, 3 in the example. Then the B-nodes from the *routing cells* form a route for B_S to B_d. In Fig. 2, the route is $B_7 \rightarrow B_5 \rightarrow B_3$ when W is zero. For the simulations presented in Section 4, W is 0. If the number of BC-nodes is small, large W can be used. For the example in Fig. 2, if W is set as $a\sqrt{2}/2$, where a is the side length of a cell, then the *routing cells* are cell 7, 5, 3; 4, 2; and 8, 6.
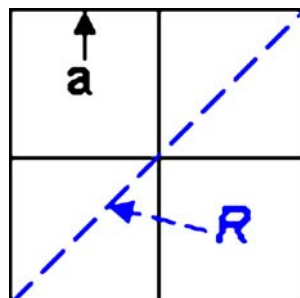
When W is larger than 0, routing among B-nodes is described in next step.

4. Route Discovery. Consider a node S wants to send a data packet to a destination node D. We first discuss the case where S is a B-node. And we will discuss the routing scheme when S is a general node in step 8. S is the starting B-node in routing. In HR protocol, the starting B-node S needs to know the current location of the destination node D. The scheme by which S obtains D's location is described in step 6 and 7. With D's location information, S knows the cell C_d in which D stays, and S knows the B-node in cell C_d is B_d. Then S determines the *routing cells* between S and B_d, and sends Route Request (RR) packets to B-nodes in *routing cells*. If W is 0, then the RR packet is just forwarded by a serial of B-node in the *routing cells*, e.g., $B_7 \rightarrow B_5 \rightarrow B_3$ as in the example. If W is larger than 0, the route is discovered as follows. The RR packet includes the following fields: Starting_B-node, Next_cells, Routing_cells, Path and Destination_cell. Starting_B-node is the B-node initiates the route discovery. Next_cells are the cells to which current B-node needs to multicast the RR packet. E.g., if $W = a\sqrt{2}/2$, the Next_cells for node B_7 is cell 4, 5, and 8. And one multicast can be realized by a single transmission. Routing_cells are the *routing cells*. Path is the route via which the RR packet traveled, and Destination_cell is the cell including the destination node. When a B-node in *routing cells* receives a RR packet, it adds itself to the Path, and determines Next_cells based on the location of itself, the location of Destination_cell, and Routing_cells. The neighbor *routing cells* in the direction towards the destination are Next_cells. Then the B-node multicasts the RR packet to B-nodes in Next_cells. When B_d receives the first RR packet, a route reply (RP) packet (with the discovered Path) is sent back via the coming path. And a route based on B-nodes is discovered. All other RR packets are discarded by B_d. Then S can start sending data packet to B_d. If S does not receive a RP from B_d for a certain time, S assumes the route discovery failed, and S will flood the RR packet to all B-nodes in the network, and finds a B-node-based route to B_d. B-node flooding is needed only when the number of B-nodes in the network is small. And flooding among a small number of B-nodes does not cause large routing overhead. Assume nodes in ad hoc networks do not move very fast. (If nodes move too fast, then besides flooding data packet, no routing protocols can successfully deliver data. Since a previously discovered route will become broken because some nodes in the route move away when data packet comes.) So most of the time, destination node D is still in cell C_d, or in nearby cells of C_d. And since B-node B_d can reach all nodes in

nearby cells, B_d can send the data packet to D directly. When D receives the packet, D will send an Ack (acknowledge) packet back to node B_d. If D is a general node, there is a unidirectional link problem here. D cannot send the Ack packet to B_d in one hop. Instead, the Ack packet is sent to B_d via flooding in cell C_d (or plus a nearby cell). If B_d does not receive the Ack for a certain time, it means D is no long in the nearby cells of C_d. Node B_d will hold the data packet (before receiving Ack) and request the updated location information of node D (described in step 6 and 7). Then B_d can send the data packet to the B-node closest to the new location of D, and that B-node can send the packet to D directly.

5. Route Repair. We will use the example in Fig. 2 to discuss the route repair in HR protocol. Assume the discovered route is $B_7 \rightarrow B_5 \rightarrow B_3$, when a B-node (e.g., B_5) in the route moves out of its cell (or fails) and no new B-node is elected, an established route becomes broken. The upstream B-node B_7 (closer to source) can detect there is no B-node in next *routing cell* if it does not overhear transmission from a B-node in the cell for a certain time after it sends the packet to the cell. Then B_7 floods route repair (RE) packets to to B-nodes in the routing cells (cell 2, 4, 6, 8 as in the example) between itself and the next B-node (e.g., B_3), to find out a path to next B-node (B_3 in the example). The RE packet includes the following fields: Repairing_B-node, Next_cell, Routing_cells, Destination_cell, Destination_ node and Path. Repairing_B-node is the B-node starting the route repair process. Next_cell is (cell 3 in the example) the next downstream routing cell after the cell without B-node (different from Next_cells in step 4). Routing_cells are the *routing cells* from Repairing_B-node to the next B-node. Path is the route via which the RE packet traveled. The RE packet is flooded among the Routing_cells. When the next B-node receives the first RE packet, a route reply (RP) packet is sent back to the Repairing_B-node B_7 via the incoming path. And a repaired route is discovered. All other RE packets are discarded by the next B-node B_3. Then the data packet can be forwarded via the repaired route.

6. Dissemination of Node Location Information. As mentioned in step 4, in HR protocol, a starting B-node S needs to know the current location of destination node D. Since nodes move around, an algorithm is needed to disseminate updated node location information. We propose an efficient dissemination scheme, and it is described in the following. If a node moves within the same cell, there is no need to update its location information. When a node moves out of its previous cell, it sends a location update packet (with its new location) to the B-node in the new cell (or the nearest B-node). The location update packet can be sent out via broadcast within a small hop count.
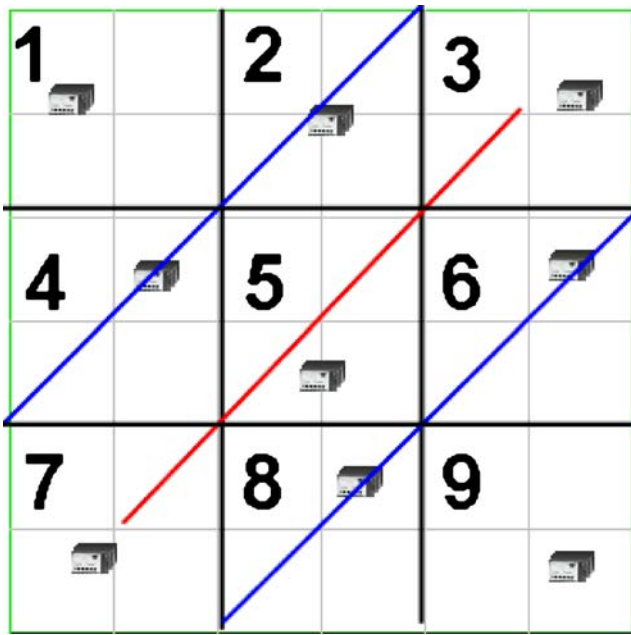
Fig. 2 Routing cells

And all B-nodes periodically send aggregated node location information to a special B-node B_0. For example, B_0 can be the commands headquarter in a battlefield. The period of updating location information should not be too large, because this will cause the location information not accurate. Also the period should not be too small, because updating the location information too often will cause large overhead. The special B-node B_0 is preferred to be a fixed B-node, or a B-node only moves within one cell. If B_0 is fixed or within one cell, the dissemination algorithm is very simple. When a starting B-node S needs to know the location of a node D, S sends a location request packet to B_0, then B_0 sends the location of D to S. Since both S and B_0 are B-nodes, they know how to communicate with each other (step 2). If B_0 also moves around, then B_0 needs to multicast its current location to all B-nodes when it moves from one cell to another. Then all B-nodes know the current location of B_0, and they can request location information from B_0. In many MANETs, it is possible to choose a static or slowly moving B-node to be B_0. And in many (like military) MANETs, it is possible and worth to deploy a static B-node as B_0.

7. There is another way to disseminate node location information. The GPS service provides the locations of all nodes to B_0, and all B-nodes request node location information from B_0. However, this approach may not be applicable to all MANETs. So for the simulations presented in Session 4, we use the scheme described in step 6.

8. Routing from General Node. If S is not a B-node, then S first needs to find out a path to a nearby B-node. S floods a Route Request (RR) packet in its cell to find the nearest B-

node. S includes its location information in the RR packet. Only nodes in the same cell as S will process and forward the RR packet to other nodes. This reduces the overhead from flooding. A node can determine if it is in the same cell as S based on the location of S and itself. The intermediate node records its identity in the *path* field of the RR packet. When the B-node B_S in the cell receives the RR packet, because B_S has large transmission range, it can send the Route Reply (RP) packet directly to node S with the *path* included. Then S can send the data packet to B_S following the *path*. And the rest is the same as in step 4. A node can discover there is no B-node in the cell if it does not receive the RP packet for a certain time after it sends out RR packet, and it will initiate a B-node election process. The B-node election process is discussed in the following.

3.2 Election of B-node

Initially, one B-node is elected in each cell if there are BC-nodes available in the cell. Since B-nodes also move around, an algorithm is needed to elect new B-node. When a B-node moves out of its current cell, it initiates the B-node election process. When a general node discovers there is no B-node in the cell (as stated in step 8 of HR protocol protocol), e.g., because the B-node failed, it initiates the B-node election process. Initially nodes know which node is the B-node in the cell. The election process works as following. The leaving B-node or the general node floods an election message to all the nodes in the cell. When a BC-node receives the election message, it broadcasts a claim message that claims it will become the B-node to all nodes in the cell. Since there is a delay from when one BC-node broadcast its claim message to when this message is heard by its neighbors, several BC-nodes may broadcast during this period. To reduce such concurrent broadcasts, a random timer is used. Each BC-node defers a random time before its B-node claim. If it hears a claim message during this random time, it then gives up its broadcast. And then one of the BC-node T becomes the new B-node in the cell, and T will start using the second address, which is the same as the cell id. Since all nodes in the cell can hear the claim message, they know that T is the new B-node. The idea is similar to the Random Competition based Clustering scheme in [18].

3.3 The new medium access control protocol

IEEE 802.11b is not an efficient Medium Access Control protocol. For HR protocol with backbone nodes, we designed a new MAC layer protocol which is more efficient than IEEE 802.11b for hybrid MANETs. The key idea is to combine

time-slotted mechanism with contention based mechanism. We call this new MAC protocol as Hybrid MAC (HMAC). In the following, we use B and G to represent B-node and general node, respectively. A large time frame is divided into three sub-frames which include G-to-G, B-to-B and B-to-G. G-to-G is a time slot assigned for the traffic between two general nodes. B-to-B and B-to-G time slots have similar meaning. The G-to-B traffic is included in G-to-G sub-frame, so there is no G-to-B sub-frame. Inside each sub-frame, the corresponding nodes use contention-based mechanism— IEEE 802.11b to decide which node should transmit packets. In B-to-B sub-frame, a B-node may piggyback another B-node the scheduling information in B-to-G sub-frame. An important parameter is the length of each sub-frame. The sub-frame length depends on the amount of traffic of each sub-type (e.g., G-to-G, B-to-B), which depends on many factors, including the number of B-nodes in the network, the number of general nodes in the network, the average distance between a general node and a B-node, and so on.

One can first use the estimated length of each sub-frame, and then use a scheme to adaptively adjust the sub-frame length of each type of traffic such that both fairness and

good throughput are achieved. The HMAC protocol is designed to minimize contention (i.e., maximize energy efficiency) and trade off fairness with throughput.

3.4 A routing example

We give a routing example by using HR protocol in Fig. 3. In the example, node 19 wants to send a packet to node 54.

1. Node 19 floods RR packet (black lines) to all the nodes in its cell. When a general node receives the RR packet, it forwards the RR packet to its neighbors. Only nodes in the same cell as node 19 will process the flooding message. The traveled *path* is recorded in the route request packet. When B-node B16 receives the RR packet, it sends a route reply packet (blue line) with the *path* directly to node 19. Then node 19 knows the *path* to node B16, in the example it is 19→38→B16, and the data packet is sent to B16.
2. Node B16 requests the location of node 54 from B_0 (node B10 in this example), and then B16 knows that node 54 is in cell 1 with the nearest B-node being B1. With $W=0$, node B16 determines the *routing cells* to be cell 16, 11, 6

Fig. 3 The routing example using HR protocol
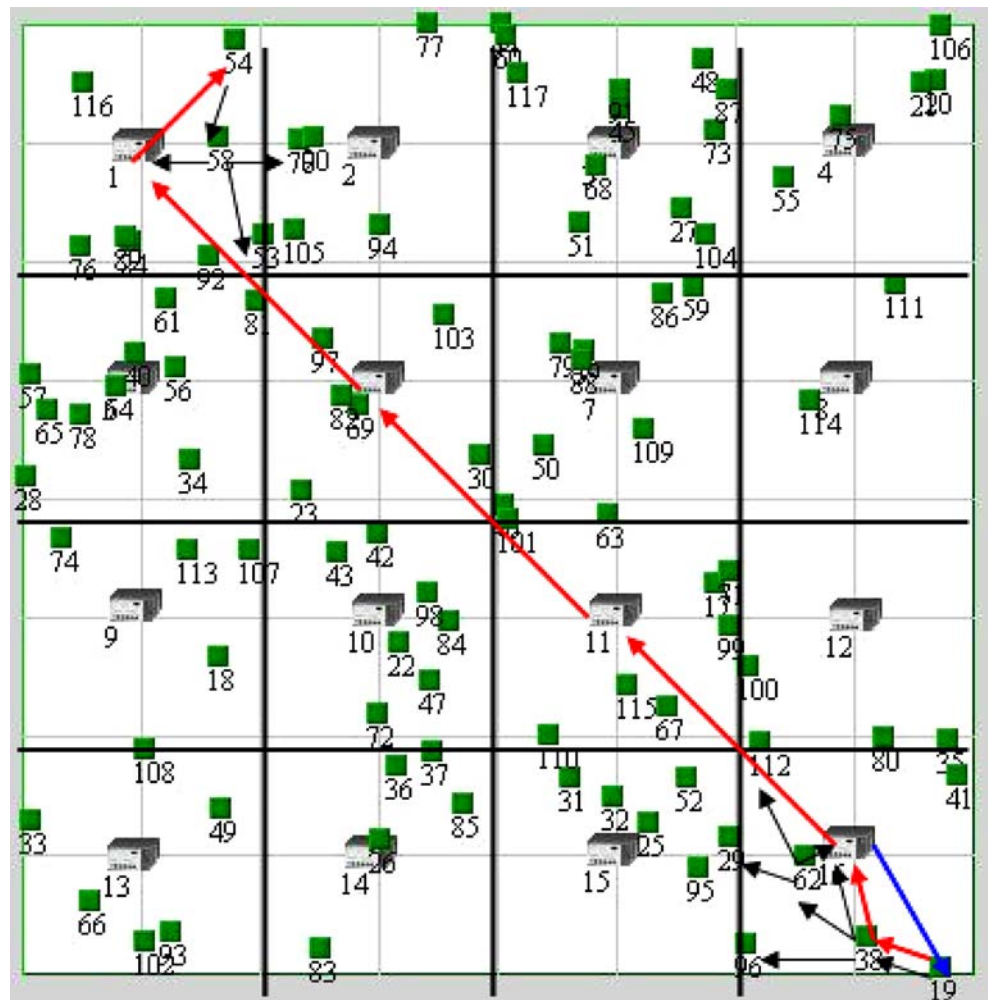
and 1. And the route is B16→B11→B6→B1 (red line). And the RR packet is sent to B1 via the route. After receiving the RR packet, B1 sends a RP packet to B16 to confirm the path availability, via the reverse path B1→B6→B11→B16.

3. Then B16 sends data packet to B1 via the route B16→B11→B6→B1.

When B1 receives the data packet, it sends the packet to node 54 by one transmission, since node 54 is still in the cell. Then node 54 sends an Ack packet (black lines) via flooding in cell 1 to node B1 to acknowledge that it receives the data packet.

As we can see from the routing example, HR provides very efficient routing. Because of the long transmission range of B-node, HR reduces the number of hops in routing. In the above example, it only takes six hops to deliver the data packet. Also, a B-node can easily determine the *routing cells* to another B-node (or cell) based on the location information and cell structure. Small number of hops and efficient B-node routing provide HR with low routing latency. Low latency is very important for routing in MANETs, because nodes in MANETs are constantly moving. Low routing latency means the intermediate nodes will not move far away from previous locations when data packet comes, and this reduces the chance of broken link. It also means the destination node is not far away from its previous location when the data packet arrives, which also reduces the routing overhead.

The routing example also shows that HR protocol has small routing overhead. It only involves a small area of flooding from the source node (or destination node) to find out the nearest B-node, plus B-node route discovery among *routing cells*. And usually the nearest B-node is very close to the source (or destination) node. If there is no B-node in the cell of the source node, then the source node S will find the path to a nearby B-node by flooding RR packet in a neighbor cell and send data packet to that B-node. The routing overhead from disseminate node location information is also not large, since the dissemination is only among B-nodes. We also want to point out that usually the number of B-nodes is small. Since the transmission range of a B-node—R is large, the side length of a cell—$a = R/2\sqrt{2}$ is also large. Then the number of cells in a fixed routing area is small. Recall that only one B-node is maintained in each cell, thus the number of B-nodes is small.

4 Performance evaluation

We have implemented HR protocol in QualNet [18], a scalable packet-level simulator with an accurate radio model. HMAC was used as the MAC protocol. For simulations presented in this Section, there are totally 20 time-slots, in which G-to-G sub-frame takes eight time-slots, B-to-B sub-frame takes eight time-slots, and B-to-G sub-frame takes four time-slots. In the following simulations, the slot allocation was fixed. We did not use any adaptive scheme to adjust the sub-frame length in the current experiments. The simulations were performed for several different network topologies. Here, we present the simulations performed by distributing 100 nodes uniformly at random in an area of $1,000 \times 1,000$ m. For HR protocol, 25 among the 100 nodes are Backbone-Capable (BC) nodes. An illustration of the simulation testbed is shown in Fig. 3. The transmission range of a B-node and a general node is 400 and 100 m, respectively. The interference range and sensing range of a general node and a BC-node are twice as the corresponding transmission range. The side length of a cell is set as $a = R/1.6$. There are 16 cells in the routing area, so there will be at most 16 B-nodes. The width of the *routing cells—W* is set to 0, i.e., *routing cells* are the cells which intercept with line *L*. Each simulation was run for 600 simulated seconds. The mobility in the environment used a random-waypoint mobility model, wherein each node randomly chooses a point in the field and moves towards it at a randomly chosen velocity. The node then pauses for a specified period at the destination before continuing the same random pattern of motion. In our simulations, the pause time was set to zero seconds, which corresponds to constant motion. We control the node mobility by varying the node velocities range. The maximum velocities vary from 0 to 50 m/s. In all simulations, B-nodes have the same set of parameters in random-waypoint mobility model as general nodes.

We set the application in the following scenario. There are several source–destination pairs in the network. The Constant Bit Rate (CBR) generators generate data traffic and sent to the destination nodes. The source sent packet of 512 b at a rate of four packets per second. We ran each simulation ten times to get an average result for each simulation configuration. We compared our HR protocol with AODV routing protocol [7], the following metrics are used to compare the routing performance.

1. Routing overhead and energy consumption. Routing overhead is the number of routing related packets (RR, RE, RP packets) for per 100 data packets. This metric is used to measure the efficiency of routing protocols. Sections 4.1 and 4.2 present the results of routing overhead comparison. The result of energy consumption comparison is also presented in Section 4.1.

2. Throughput and delay. These two metrics are used to measure the effectiveness of routing protocols. Sections 4.3 and 4.4 present the results of throughput and delay comparison.
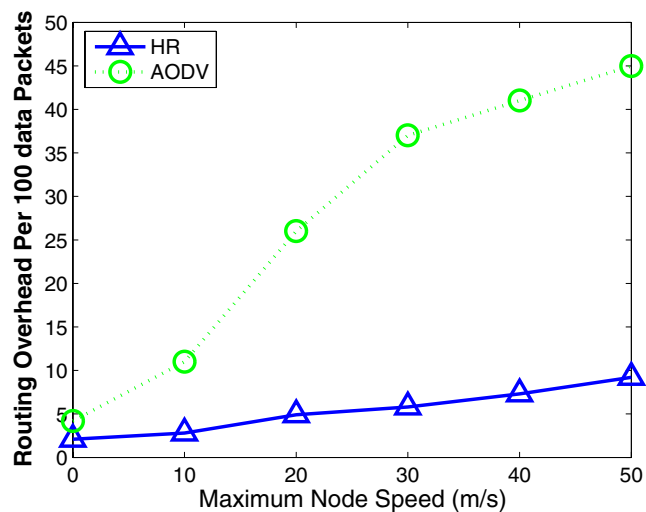
Fig. 4 Routing overhead under different mobility

3. Scalability. Scalability is a very important metric for MANET routing. The results of scalability test are reported in Section 4.5.

In addition, we evaluate the performance of HMAC in Section 4.6. Then in Sections 4.7 and 4.8, we discuss the probability of having a B-node in a cell and how BC-node density affects routing overhead. In the following tests, the routing overhead of HR includes the overhead of disseminating node location information.

4.1 Routing overhead under different mobility

First we compared the routing performance of HR with AODV for different node mobility. Figure 4 is the average routing overhead for send 100 data packets when node maximum speed varies from 0 to 50 m/s. The routing overhead of both HR protocol and AODV increases as the node speed increases. Higher mobility causes more broken links, and thus increases routing overhead. However, the overhead of HR is much smaller than AODV. And we can see that the routing overhead of HR does not increase much as the node speed increases, while the overhead of AODV increases very fast as the node speed increases. When the maximum node speed is 0 m/s (no mobility), the overhead of AODV is close to that of HR. However, the overhead of AODV is about six times as that of HR when the maximum node speed is 50 m/s. Since B-nodes have large transmission range, the number of hops in HR is small, and the routing latency in HR is low. Low routing latency means the time between route discovery and sending data packet is small, thus the intermediate nodes are not far away from their previous locations when the data packet comes, i.e., less broken links. Hence, for high node mobility, HR incurs much less routing overhead than AODV.

We also measured the total energy consumption of HR and AODV for different node mobility. The result is plotted in Fig. 5, where the total energy consumption in the network for delivering every 100 data packets. Figure 5 shows that AODV consumes much more energy than HR for all the tested node speeds. The reason is obvious—larger overhead means more energy consumption.

4.2 Routing overhead vs. transmission range

We then studied the routing overhead when node transmission range varies. Here we only change the transmission range of general nodes. The transmission range of B-node does not change, and it is always 400 m. Figure 6 shows the effect of varying the transmission range on the routing overhead. The overhead is the average of sending 100 data packets. We compared the routing overhead of HR and AODV for different node transmission range. The transmission range of a general node varies from 40 to 200 m, with an increase of 20 m. Figure 6 shows that the overhead of HR is much smaller than that of AODV. When the transmission range is larger than 140 m, the overhead of HR is very small, less than five packets for 100 data packets. This is because when the transmission range is large enough, source node and destination node can communicate with nearby B-node directly. That is, source node can send data packets to nearby B-node directly, and destination node can send the Ack packet directly to the nearby B-node. So there is no need to flood route request or Ack packet, hence reduces the routing overhead a lot.

4.3 Throughput under different traffic load

To evaluate the effectiveness of HR protocol protocol, we measured the throughput and delay of HR under different traffic load, and we compared the performance with AODV.
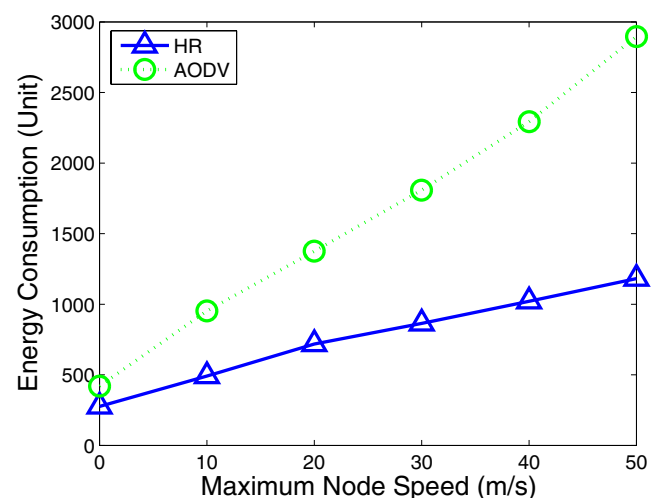

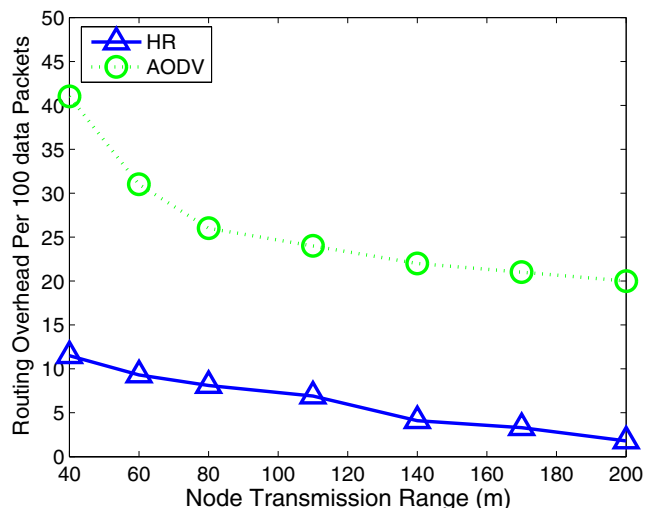Fig. 5 Energy consumption under different mobility

Fig. 6 Routing overhead vs transmission range

In this subsection, we present the result of throughput comparison in Fig. 7. In the simulation, we vary the traffic load from 50 to 300 kbits/s. The maximum node speed is 20 m/s. Figure 7 shows that under AODV, the network starts saturation when the traffic load is heavy, and the throughput actually decreases when traffic load is larger than 250 kbits/ s. This is because AODV uses flooding to discover route, and flooding incurs large routing overhead. When the network traffic is heavy, congestion happens and packets are dropped in the network, and cause the throughput decreases.

4.4 Delay comparison

The delays of the HR and AODV under different traffic load are plotted in Fig. 8. As we can see, the delay of AODV is close to HR when traffic load is light, i.e., less than 150 kbits/s. Under light traffic, the network does not
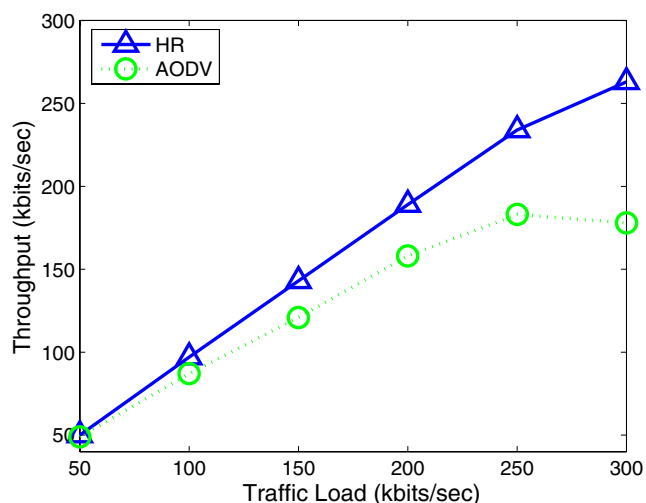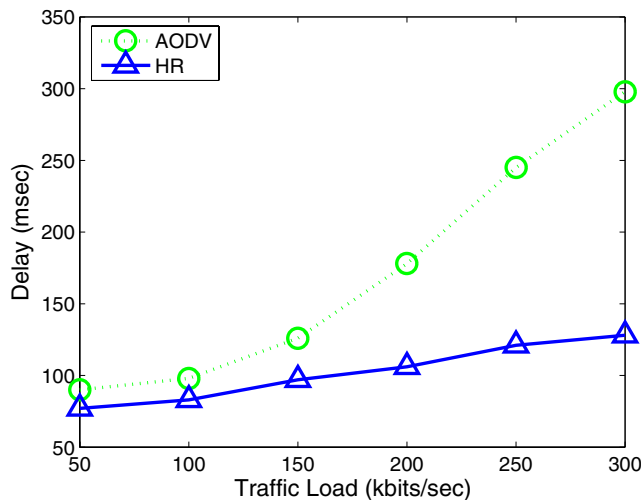

Fig. 8 Delay comparison

have many congestions or packet drops, the packet delays of both protocols are not large, so they are close to each other. The routing overhead of AODV is larger than HR, and more packets in the network means larger queuing delay. When the traffic load becomes heavy, the delay of AODV increases very fast, which means lots of congestions and packet drops are caused in the network, and a large number of packets incur long-time delay or are dropped. On the other hand, HR has much smaller routing overhead, there is little congestion and packet drops even under heavy traffic load. This is why the delay under HR increases slowly as the network traffic becomes heavy.

4.5 Scalability

Scalability is a very important aspect of MANET routing protocol design. A good routing protocol should scale well in large networks. That is, the routing performance should
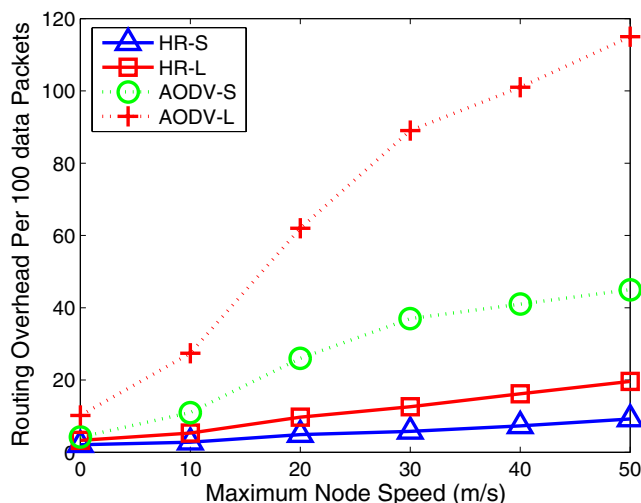

Fig. 7 Throughput comparison

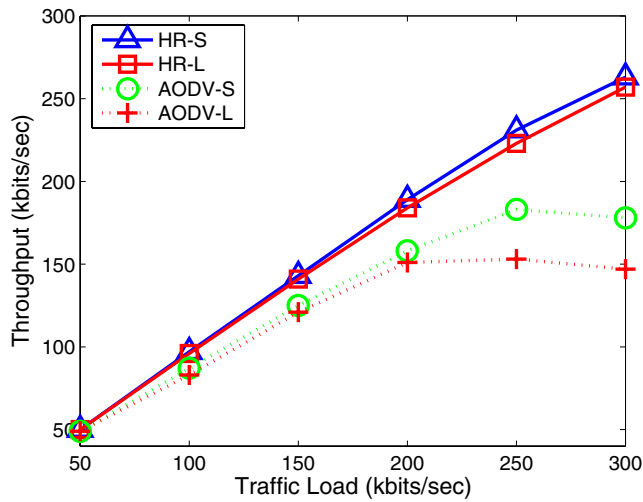Fig. 9 Scalability—routing overhead

Fig. 10 Scalability—throughput

Fig. 12 The delay comparison

deteriorate reasonably when network size increases. In order to test the scalability of HR, we implemented a large testbed with 400 nodes distributed uniformly at random in an area of 2,000×2,000 m, where 100 nodes are Backbone-Capable nodes. The large testbed is four times the small testbed (in terms of node number and routing area) used in previous tests. We compare the performance of HR and AODV in the large testbed.

Figure 9 reports the routing overheads of HR and AODV for different mobility in large testbed, which are labeled as HR-L, AODV-L. For comparison, in Fig. 9, we also plot the routing overheads of HR and AODV in small testbed, which are labeled as HR-S, AODV-S. Figure 9 shows that the routing overhead of HR-L (in large testbed) is about twice as the overhead of HR-S, and HR-L is not large even under high mobility. But the routing overhead of AODV-L is much larger than AODV-S, especially under high mobility. When the network size increases, the flooding
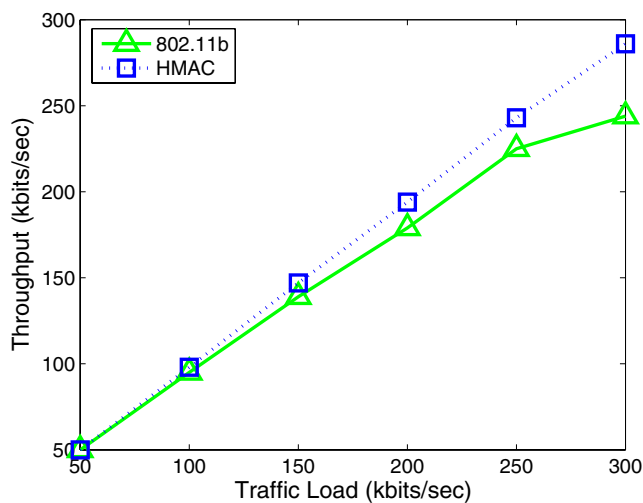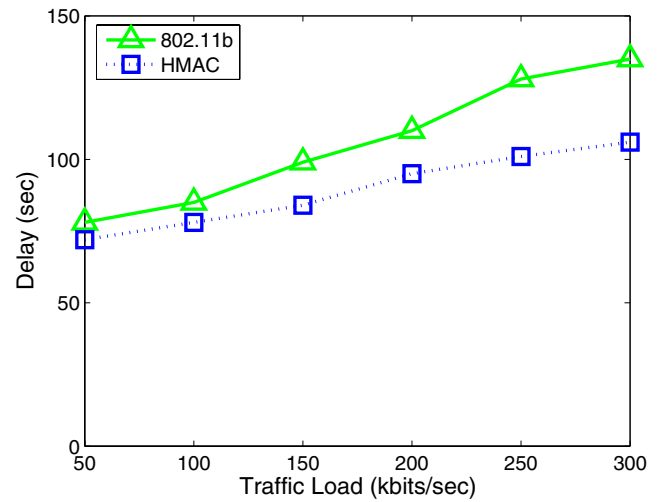
area of AODV route discovery also increases, and much more nodes participate in route discovery, so the overhead of AODV-L is much larger than AODV-S. Also, in large network, on average AODV needs more intermediate nodes to forward a packet. If any intermediate node fails or moves away, the route has to be repaired or re-discovered, and this causes large routing overhead. For the same source–destination pair, HR needs much less intermediate nodes than AODV because HR uses B-nodes to route packets, which have long transmission range. Thus, the chance of a broken route is much smaller in HR. Also HR has low routing latency as mentioned in Section 4.1. Hence the routing overhead of HR in large networks is still small.

We also measured the throughputs of HR and AODV in the large testbed, and the results are reported in Fig. 10. For comparison, the throughputs of HR and AODV in the small testbed are also plotted in Fig. 10. The labels have the same meaning as in Fig. 9. As we can see, the throughput for
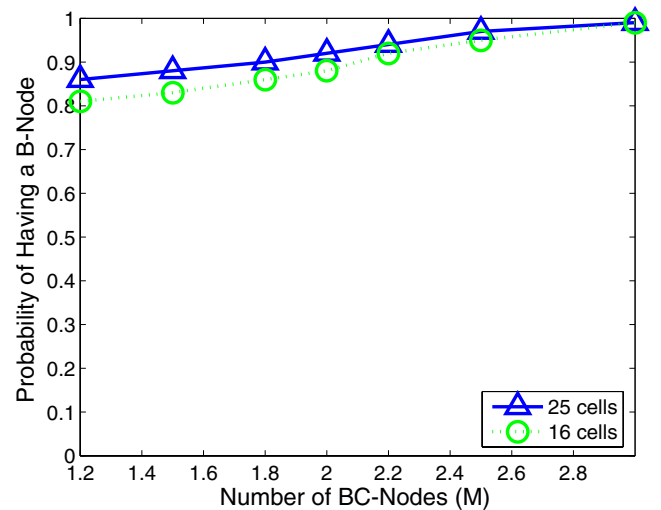

Fig. 11 The throughput performance

Fig. 13 The probability of having B-nodes

HR-L is almost the same as HR-S. That is, HR is also very effective in large networks. However, the performance of AODV becomes worse in the large network than in the small network. The throughput of AODV-L drops when traffic load is larger than 200 kbits/s.

4.6 The performance of HMAC

We studied the performance of the Hybrid MAC protocol. We compared the throughput and delay of HR protocol using HMAC and IEEE 802.11b, and the results are reported in Figs. 11 and 12. Figures 11 and 12 show that HMAC provides higher throughput and incurs smaller delay than IEEE 802.11b.

4.7 The probability of having B-nodes

To ensure the good performance of HR protocol, it is important to have B-nodes in most cells. If there is no B-node in a cell, then B-node flooding (during route discovery) or small-area flooding (during route repair) will have to be used, which increase routing overhead. In [8], Du proved that there is a high probability that each cell has at lease one B-node with a reasonable number of BC-nodes (about 1.5 times the number of cells) in the network. We also run simulations to measure the probability of having a B-node in a cell. Figure 13 lists the simulation results of two networks, one with 16 cells and the other with 25 cells. The x-axis is the number of BC-nodes in the network, in terms of the number of cells M. Figure 13 shows that the probability is very high for reasonable number of BC-nodes, and it increases as the number of BC-nodes increases. Please refer to [8] for more details about this issue.

4.8 Performance for different BC-node densities

We also use simulation to evaluate the performance of HR protocol for different BC-node densities. In the simulation, the node maximum speed (of both HR and AODV) is 20 m/s. The traffic load in the experiment is 150 kbits/s. The number of BC-nodes in HR varies from 5 to 35, with an increase of 5. Recall that there are 16 cells in the simulation network. We compare the routing overhead of HR and AODV for different densities of BC-nodes. The result is reported in Fig. 12, which is the average routing overhead for sending 100 data packets. Figure 14 shows that HR has larger routing overhead than ZRP when the density of BC-nodes is low. The routing overhead of HR decreases quickly as the number of BC-nodes increases. We have explained this issue in Section 4.7.

Based on our HR protocol, we propose a novel concept for mobile ad hoc network design and deployment—Active Routing. In Active Routing, a small number of backbone
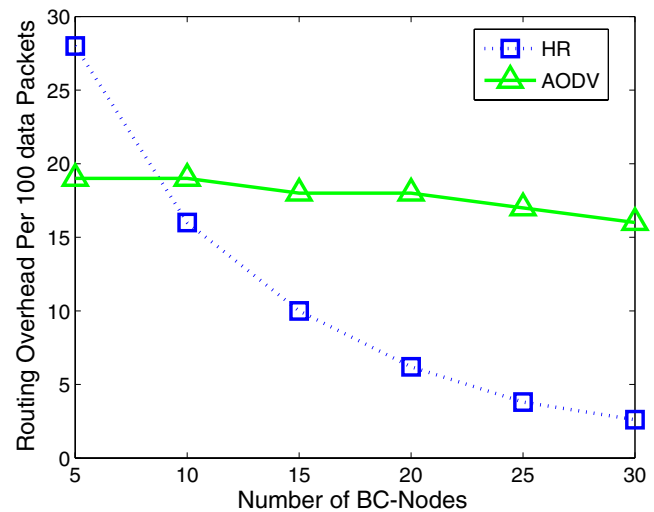

Fig. 14 Overhead for different BC-node densities

nodes are placed at some desired locations to aid routing. Active Routing can significantly improve routing performance. In traditional MANET routing protocols, researchers usually assume there is no control of the location of any node. However, in certain MANETs, it is possible to control the location (or mobility) of some nodes. For example, in a large military MANET, it is possible and worth to deploy a small number of relay nodes at desired locations mainly for communication purpose. Active routing is a novel concept for the design and deployment of mobile ad hoc networks, and this system design concept is very useful for many MANETs.

5 Conclusion

This paper presents Hybrid Routing—a novel routing protocol that takes advantage of the powerful nodes in many realistic mobile ad hoc networks. Different types of nodes have different transmission power (range), data rate, processing capability, reliability and security. HR protocol utilizes the more powerful nodes as backbone nodes (B-nodes). The routing area is divided into multiple cells, and one B-node is maintained in each cell if possible. Based on location information and cell structure, routing among B-nodes is very efficient and simple. An efficient algorithm is provided to disseminate node location information among all B-nodes. A source discovers a route to a destination in an on-demand way, and most of the routing activities are carried out by powerful B-nodes. This reduces the number of routing hops and makes the routing more efficient and reliable, since B-nodes have large data rate and transmission range, and are more reliable. We also proposed a novel HMAC protocol that can improve the efficiency of medium access in hybrid MANETs. Extensive simulations demon-

strate that HR has very good performance, and performs much better than AODV in terms of reliability, scalability, route discovery latency, routing overhead as well as packet delay and throughput. Our simulations also show that HMAC is more efficient than IEEE 802.11b.

Reference

1. Gupta P, Kumar PR (2000) The capacity of wireless networks. IEEE Trans Inf Theory IT-46(2):388–404 (March)
2. Xu K, Hong X, Gerla M (2002) An ad hoc network with mobile backbones. In: Proc. of IEEE ICC 2002, New York, NY (Apr.)
3. Ko YB, Vaidya NH (1998) Location-aided Routing (LAR) in mobile ad hoc networks. In: Proc of ACM/IEEE MobiCom, pp 66–75
4. Joa-Ng M, Lu I-T (1999) Peer-to-peer zone-based two-level link state routing for mobile ad hoc networks. IEEE JSAC 17 (8):1415–1425 (Aug.)
5. Karp B, Kung HT (2000) GPSR: greedy perimeter stateless routing for wireless networks. In: Proc of MobiCom
6. Ye Z, Krishnamurthy SV, Tripathi SK (2004) A routing framework for providing robustness to node failures in mobile ad hoc networks. Ad Hoc Networks 2(1):87–107
7. Perkins CE, Royer EM Ad-hoc on demand distance vector routing. In: Proc of IEEE WMCSA '99, New Orleans, LA, pp 90–100
8. Du X (2004) QoS routing based on multi-class nodes for mobile ad hoc networks. Journal of Ad Hoc Networks 2/3:241–254 (July)
9. Liao WH, Tseng YC, Sheu JP (2001) GRID: a fully location-aware routing protocol for mobile ad hoc networks. Telecommunication Systems 18(1–3):37–60
10. Basagni S et al (1998) A distance routing effect algorithm for mobility (DREAM). In: Proc of ACM/IEEE MobiCom, pp 76–84
11. Lee S-J, Su W, Gerla M (2001) Wireless ad hoc multicast routing with mobility prediction. Mobile Networks and Applications 6(4): 351–360 (Aug.)
12. Haas ZJ et al (2001) The performance of query control schemes for the zone routing protocol. IEEE/ACM Trans Netw 9:427–438 (Aug.)
13. Sivakumar R et al (1999) CEDAR: a core-extraction distributed ad hoc routing algorithm. In: Proc of INFOCOM
14. Ramasubramanian V et al (2002) Providing a bidirectional abstraction for unidirectional ad hoc networks. In: Proc of INFOCOM
15. Bao L, Garcia-Luna-Aceves JJ Link state routing in networks with unidirectional links. In: Proc of IEEE ICCCN 99. Boston, MA
16. Sinha P, Krishnamurthy S (2000) Scalable unidirectional routing with zone routing protocol (ZRP) extenstions for mobile ad hoc networks. In: Proc of WCNC, Chicago (Sept.)
17. Gerla M, Hong X, Pei G (2000) Landmark routing for large ad hoc wireless networks. In: Proc of IEEE Globecom 2000, San Francisco, CA, (Nov.)
18. QualNet Network Simulator, The scalable network technology, http://www.qualnet.com

Xiaojiang (James) Du is an assistant professor in Dept. of Computer Science, North Dakota State University. Dr. Du received his B.E. degree from Tsinghua University, Beijing, China in 1996, and his M.S. and Ph.D. degrees from University of Maryland, College Park in 2002 and 2003, respectively, all in Electrical Engineering. His research interests are wireless sensor networks, mobile ad hoc networks, wireless networks, computer networks, network security and network management. Dr. Du is an associated editor for John Wiley Journal of Wireless Communication and Mobile Computing. He is the program chair of Computer and Network Security Symposium of IEEE International Wireless Communication and Mobile Computing Conference (IWCCC) 2006.

Dapeng Wu received B.E. in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, M.E. in Electrical Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and Ph.D. in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 2003. Since August 2003, he has been with Electrical and Computer Engineering Department at University of Florida, Gainesville, FL, as an Assistant Professor. His research interests are in the areas of networking, communications, multimedia, signal processing, and information and network security. Dr. Wu received the IEEE Circuits and Systems for Video Technology (CSVT) Transactions Best Paper Award for Year 2001.