# Efficient On-Demand Routing for Mobile Ad Hoc Wireless Access Networks

Joo-Han Song, *Student, IEEE*, Vincent W. S. Wong, *Member, IEEE*, and Victor C. M. Leung, *Fellow, IEEE*

***Abstract*—In this paper, we consider a mobile ad hoc wireless access network in which mobile nodes can access the Internet via one or more stationary gateway nodes. Mobile nodes outside the transmission range of the gateway can continue to communicate with the gateway via their neighboring nodes over multihop paths. On-demand routing schemes are appealing because of their low routing overhead in bandwidth restricted mobile ad hoc networks, however, their routing control overhead increases exponentially with node density in a given geographic area. To control the overhead of on-demand routing without sacrificing performance, we present a novel extension of the ad hoc on-demand distance vector (AODV) routing protocol, called LB-AODV, which incorporates the concept of load-balancing (LB). Simulation results show that as traffic increases, our proposed LB-AODV routing protocol has a significantly higher packet delivery fraction, a lower end-to-end delay and a reduced routing overhead when compared with both AODV and gossip-based routing protocols.**

***Index Terms*—Load-balancing, mobile ad hoc network (MANET), on-demand routing protocol, wireless access network.**

## I. Introduction

A MOBILE ad hoc network (MANET) [1] consists of a set of wireless mobile nodes communicating with each other without any centralized control or fixed network infrastructure. MANETs have been evolving to serve a growing number of applications that rely on multihop wireless infrastructures that can be deployed quickly. The potential applications include emergency disaster relief, battlefield command and control, mine site operations, and wireless classrooms or meeting rooms in which participants wish to share information or to acquire data.

Today, advances in wireless technologies such as IEEE 802.11 [2], Bluetooth [3], and third-generation cellular have led to a proliferation of mobile devices. The number of mobile Internet devices is expected to reach a billion in the near future [4] and exceed the number of stationary nodes. Therefore, we expect MANETs to be interconnected to the Internet in many applications. In this paper, we consider a mobile networking environment in which mobile hosts can access the Internet directly via one or more gateways or access points, or indirectly

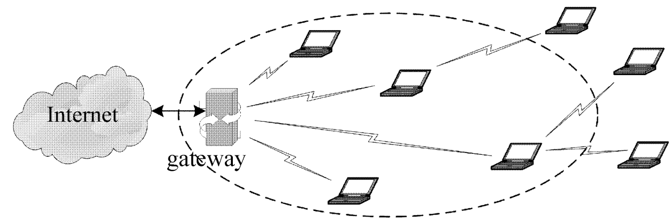



Fig. 1. Mobile ad hoc wireless access network.

via other mobile hosts. This is referred as a *mobile ad hoc wireless access network* or *wireless mesh network*. Mobile hosts that are near the gateway can communicate directly with the gateway via single-hop connections. However, mobile hosts that are outside the transmission range of the gateway have to use multihop connections that rely on the neighboring mobile nodes to relay their packets (see Fig. 1). In addition to providing Internet access, this network configuration may also serve other practical scenarios; e.g., the gateways may represent nodes that host special services such as domain name service (DNS) accessed by other nodes in the MANET.

Consider the effects of the total number of mobile nodes in a MANET over a given coverage area. Even though increasing the total number of mobile nodes tends to reduce the effective bandwidth available at individual nodes due to increased competition for bandwidth, it also increases the connectivity of the network, which may be important as node mobility increases. Results in [5]–[7] show that when the number of nodes is small, the network may not be fully connected, in that some nodes may not be able to send packets to certain destinations. Network connectivity can be increased by simply increasing the total number of mobile nodes in the network. However, when the number of nodes and the traffic load increase, contention and packet collisions between neighboring nodes also increase exponentially [8]. Therefore, there is a tradeoff between maintaining the full network connectivity and minimizing bandwidth contention in a MANET.

Since on-demand routing protocols (e.g., ad hoc on-demand distance vector (AODV) [9] and dynamic source routing (DSR) [10]) use the flooding method to find a route to the destination, the number of rebroadcasts of route request (RREQ) packet is proportional to the number of nodes. Therefore, the routing control overhead increases with the total number of nodes. Consider the situation where mobile node $A$ broadcasts an RREQ message to $n$ neighboring nodes. The $n$ neighbors may rebroadcast this RREQ message to their respective neighbors. Packet collisions may occur over the wireless medium, resulting in congestion and possible loss of routing control packets. Furthermore, the source node may attempt to recover from loss of routing

Manuscript received October 1, 2003; revised March 15, 2004. This work was supported in part by a University of British Columbia Graduate Fellowship, and in part by the Canadian Natural Sciences and Engineering Research Council under Grant RGPIN 262604-03 and Grant 44286-00. This paper was presented in part at the IEEE Vehicular Technology Conference Spring'03, Jeju, Korea, April 2003, and in part at the IEEE Globecom'03, San Francisco, CA, December 2003.

The authors are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: joohans@ece.ubc.ca; vincentw@ece.ubc.ca; vleung@ece.ubc.ca).

Digital Object Identifier 10.1109/JSAC.2004.829334

control packets by initiating another route discovery process, which further increases the amount of control traffic in the network [11]–[13]. In order to maintain a high packet delivery fraction and a low end-to-end delay for packet transmissions over a MANET, it is important to reduce the amount of routing control traffic [14], [15].

The above discussion motivates us to design an efficient routing mechanism, which can find a route to the gateway with a controlled amount of routing overhead. In this paper, we propose a novel extension of the AODV routing protocol for mobile ad hoc wireless access networks, which applies the concept of load-balancing to limit the amount of routing control packets. In our proposed load-balancing scheme, AODV route selection is regulated by a distributed grouping mechanism, which divides the mobile nodes logically into different groups to reduce and distribute routing traffic over the network. Load-balancing is accomplished by balancing the number of source nodes among the groups, a process that can be controlled and updated by the gateway(s).

The rest of this paper is organized as follows. Our proposed extension of the AODV routing protocol with load-balancing (LB-AODV) is described in Section II. Simulation results for performance comparisons are presented in Section III. Conclusions are given in Section IV.

## II. Ad Hoc On-Demand Distance Vector Routing Protocol With Load-Balancing (LB-AODV)

In this section, we begin by describing the rationale and operation of our load-balancing mechanism based on grouping of mobile nodes. The operation of the proposed LB-AODV routing protocol is explained in the Sections II-B and II-C. It is followed by a discussion on the selection of the total number of groups in Section II-D. Balance index update procedures are introduced in Section II-E. Finally, we compare the route discovery processes among AODV, gossip-based, and LB-AODV routing protocols in Section II-F.

We shall initially consider a mobile ad hoc wireless access network with a single gateway. The following terminologies will be used in this paper: A *source node* is a mobile node with data packets to send toward the gateway. A *common node* is a mobile node that does not have data to send and does not belong to any particular group. An *active node* is a mobile node that has valid route(s) to the gateway and is currently being used to forward packets toward the gateway.

### *A. Load-Balancing Mechanism With Grouping*

We propose a load-balancing mechanism based on the concept of grouping. It reduces the number of unnecessary retransmissions of routing messages and prevents network congestion by separating source nodes into different *groups* and allowing source nodes to relay only packets generated by their own group members and common nodes [16].

The basic idea of our grouping mechanism is to partition all mobile nodes into several logical divisions such as *A*, *B*, *C*, *D*, and *E*, as shown in the example in Fig. 2. All common nodes belong to the division *E* in this example, and they are allowed to relay packets from any groups toward the gateway. On the other

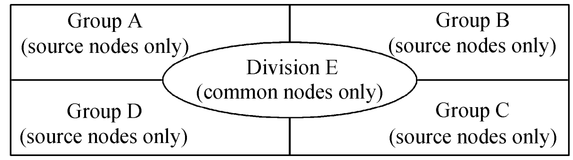



Fig. 2. Example of logical partitioning of mobile nodes.

hand, a source node, which belongs to one of the groups *A*, *B*, *C*, and *D* in this example, is not allowed to relay packets from other than its own group. For example, packets generated by any members of group *A* can be relayed only by other source nodes of group *A* and common nodes belonging to division *E*.

In the route discovery process, an RREQ message is only forwarded to the common nodes and those nodes that belong to the same group. Thus, the amount of control traffic can be reduced. The determination of the number of groups is an important consideration in the operation of the LB-AODV routing protocol, and is discussed in detail in Section II-D.

By dividing source nodes into several groups, the packet relaying responsibility and the traffic load can be distributed among different groups. The proposed load-balancing mechanism aims at maximizing the balance index $B$, which is defined as [17]

$$B = \frac{\left[\sum_{i=1}^{G} f_i\right]^2}{G \sum_{i=1}^{G} f_i^2} \tag{1}$$

where $f_i$ denotes the total number of source nodes of group $i$, and $G$ denotes the total number of groups. Given the number of groups $G$, the balance index converges to 1 when the total number of source nodes of each group approaches equality, while it approaches $1/G$ when all source nodes of the network are assigned to the same group. In our LB-AODV routing protocol, the *state information* is a $(G + 1)$—tuple in the form of $\langle \mathit{group\ number}, f_1, f_2, \ldots, f_G \rangle$. This information is maintained at all active mobile nodes.

The idea of grouping nodes in LB-AODV is similar to the concept of *routing zone* in the zone routing protocol (ZRP) [18]. Both routing protocols send queries only to selected nodes in the network during the route discovery process. However, the design goals for these two protocols are different. ZRP targets toward self-organized mobile ad hoc networks, while LB-AODV targets toward mobile ad hoc wireless access networks in which mobile nodes can access the Internet via stationary gateway node(s). In addition, the zone partitioning in ZRP is physical in which nodes within certain number of hops are being grouped together. On the other hand, LB-AODV partitions mobile nodes into several logical divisions in order to maximize the balance index. Furthermore, ZRP belongs to a family of hybrid proactive/on-demand routing protocols, whereas LB-AODV is a purely on-demand routing protocol.

### *B. Load-Balancing Route Decision Process*

Using the load-balancing route discovery process, we can dynamically minimize the variance of the total number of source nodes between groups. The flow charts for the route selection

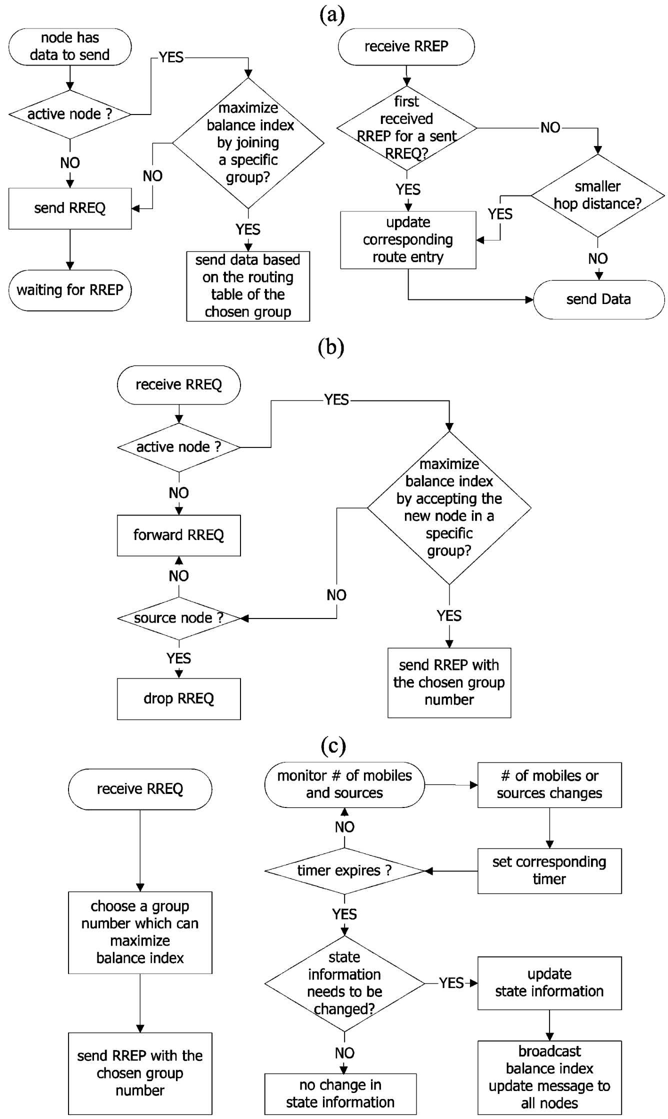



Fig. 3. Processes in mobile nodes and gateway. (a) Source node. (b) Intermediate node. (c) Gateway node.

process are shown in Fig. 3. A group number is assigned to each source node that initiates the route discovery process.

When a node has data to send but does not know a route to the gateway, the new source node initiates the route discovery process by broadcasting an RREQ message to its neighboring nodes [see Fig. 3(a)]. When an intermediate node receives the RREQ packet, it processes this message according to its state information. An intermediate node that is not an active node will simply broadcast this RREQ message to its neighbors. On the other hand, if the intermediate node is an active node, it will calculate the balance index $B$ based on the state information stored in its cache. If the balance index $B$ can be maximized by accepting this new source node into one of its serving groups, then this intermediate node will send a route reply (RREP) message to the source node. This RREP message includes information about which group this particular source node has been assigned to. The flow chart for the operations of an intermediate node is shown in Fig. 3(b). Since the active intermediate node can assign different groups to the source node according to its state information, it needs to maintain different route entries to the gateway for different groups it is currently serving.

Similarly, when the gateway node receives an RREQ message, it will assign a group number to the new source node. The group number is chosen such that the balance index $B$ is maximized. The gateway then sends an RREP message to the source node. The flow chart for the operations of the gateway node is shown in Fig. 3(c). When the source node receives the RREP message, it will begin sending data packets to the gateway immediately via the node from which it received the RREP message.

Due to transient balance index mismatch, it is possible that a source node may receive multiple RREP packets from several nodes with different group numbers. In this case, the source node will compare the hop count field in those RREP packets and select the one with the smallest hop count field value. The group number in the chosen RREP packet will then be used.

### C. *Load-Balancing Route Maintenance Process*

When a source node detects a link breakage via the route error (RERR) message, it will reinitiate the route discovery by sending an RREQ message with its group number toward the gateway. Those intermediate source nodes that do not belong to this particular group will simply drop the RREQ message. When either an active node (which has a routing cache for this group) or another source node (which belongs to the same group) receives the RREQ message, it will send an RREP message to the source node. The above procedures limit the amount of routing overhead. Note that the balance index remains unchanged after the route discovery process. This is because the new route is still part of the original group.

Due to the topology changes brought about by node mobility, it is possible that the RREQ message may not reach the gateway via the routes in a particular group. We resolve this issue as follows: If the source node has not received any RREP message after a certain period of time, it will reinitiate the route discovery process, as if it was a new source node, by sending another RREQ message without the group number. As long as there exists a route to the gateway, the source node will eventually join another group.

When an active intermediate node becomes a new source node, it first checks the state information stored in its cache. If the balance index can be maximized via one of its serving groups (e.g., group $x$), then this new source node will send data packets to the gateway using group number $x$. Otherwise, this new source node will initiate the route discovery process by broadcasting the RREQ message to its neighboring nodes to find a route that can maximize the balance index.

We assume that soft state information is maintained in the routing cache in each active node. That is, each routing entry has an associated timer. When an intermediate active node or gateway has not received data packets corresponding to a particular entry for a certain period of time, that routing entry and its group number will be deleted.

### D. *Determining the Number of Groups*

The determination of the number of groups is critical for the efficiency of the LB-AODV routing protocol. The number of groups is chosen as a tradeoff between the network connectivity and the amount of routing control overhead. To determine the number of groups, the gateway has to obtain the following information: the number of source nodes, the number of mobile nodes, and the size of network. Different methods exist for the estimation of these parameters. In this paper, we assume that the gateway can estimate the number of source nodes by monitoring the source address field in the packet header from the packets it received. Assuming that network authorization and authentication are required for the mobile nodes to communicate with the gateway, the gateway can estimate the total number of mobile nodes. If the size of network is not known in advance, the gateway has to estimate the size of the network based on the hop count information from the packets it received. Due to estimation errors, the size of network and the corresponding group number may not always be correct. To measure the percentage change of throughput due to estimation error, we perform a sensitivity analysis in Section III-C.

It has been shown [5] that for normal MANET scenarios, the best performance can be achieved when the average number of neighbors is around seven. In this paper, we define the *optimal number of mobile nodes* $R$ in a MANET *topology*[1] as the number of mobile nodes that results in the average number of neighbors being around seven. If the gateway can estimate the size of the network correctly, it can calculate the value of $R$ by assuming that all the nodes are evenly distributed in the network.

We aim to minimize the difference between the optimal number of mobile nodes $R$ and the variable $T$, which is defined as *the total number of mobile nodes that can relay packets generated by each group*. The rationale is that the optimal number $R$ gives the best performance without decreasing the network connectivity in a given network.

Given the number of source nodes $S$ and the number of mobile nodes $M$, the total number of common nodes is equal to $M - S$. Given the total number of groups $G$ and assuming that each group has the same number of source nodes, the number of source nodes that belong to each group is $S/G$. The total number of mobile nodes $T$ that relay packets generated by each group is given by $M - S + (S/G)$. Therefore, the gateway chooses the number of groups $G$ such that the absolute difference between $T$ and $R$ is minimized

$$\begin{aligned} G &= \arg\min_{g \in \{1,2,\cdots,S\}} |T - R| \\ &= \arg\min_{g \in \{1,2,\cdots,S\}} \left| M - S + \left(\frac{S}{g}\right) - R \right|. \end{aligned} \quad (2)$$

The number of groups $G$ is the function of the value of $M$, $S$, and $R$, as shown in (2).

1) If $M \geq R + S$, then $G = S$. Because the number of common nodes, $M - S$, is greater than or equal to the optimal number of mobile nodes $R$ in a given topology, a single source node per group will minimize $|T - R|$.
2) If $R < M < R + S$, then $G$ is equal to one of $\lfloor S/(R + S - M) \rfloor$ or $\lceil S/(R+S-M) \rceil$, which minimizes $|T - R|$.
3) If $M \leq R$, then $G = 1$. Since the total number of mobile nodes $M$ is less than the minimum required number of mobile nodes $R$, all the mobile nodes have to join the same group to minimize $|T - R|$. In this case, our LB-AODV routing protocol is identical to the original AODV.

[1]Topology here refers to a specific grid size over which mobile nodes are uniformly distributed.

Consider the following example: The network topology is $1500 \times 300$ ($\mathrm{m}^2$). The location of all the nodes is uniformly distributed in the network. The module *setdest* in network simulator (*ns2*) [19] can be used to calculate $R$ such that the average number of neighbors is around seven. Based on this calculation, $R$ is equal to 30. In this case, if the number of source nodes $S$ is 25 and the number of mobile nodes $M$ is 50, then from (2), the number of groups $G$ is equal to 5.

### E. Balance Index Update

Our proposed LB-AODV can also support dynamic changes in the number of groups as the number of mobile nodes changes due to either join or leave operations. We assume that the gateway monitors the total number of mobile nodes $M$ and the number of source nodes $S$ periodically. Whenever the optimal number of groups $G$ or the number of source nodes has changed [from (2)], the state information needs to be updated based on (1). The gateway then broadcasts an advertisement message to all the nodes to update the state information. The update information includes: 1) the number of source nodes in each group $\langle f_1, f_2, \ldots, f_G \rangle$ and 2) the addresses of those source nodes that have been reassigned to different groups and their newly assigned group numbers. For those source nodes that have been assigned new group numbers, they will reinitiate the route discovery process again by including the new group number in the subsequent RREQ messages.

### F. Comparison of Route Discovery Processes

In this section, we describe the differences in the route discovery procedures among the AODV, gossip-based, LB-AODV routing protocols. Consider a MANET with a large number of mobile nodes. We assume that each mobile node has $n$ neighbors within its transmission range, and none of the mobile nodes has a route entry to the requested destination. Suppose a source node $S$ sends an RREQ message to its neighboring nodes. In all three routing protocols, we consider that when a neighboring node first receives an RREQ packet, the node will either broadcast the RREQ packet to its neighbors with probability $p$ or discard it with probability $1 - p$, where $0 \leq p \leq 1$. If the node receives the same RREQ packet again, the packet will be discarded. Different protocols differ in the possible values of $p$.

In the AODV routing protocol, when a mobile node first receives an RREQ packet and does not have a route entry to the requested destination, it will always broadcast the RREQ packet to its neighbors. Therefore, an RREQ packet will be broadcasted over more than one hop with probability $p = 1$. However, the number of RREQ packets being broadcasted is proportional to the number of nodes, and cannot be controlled or regulated.

Consider the basic gossip-based routing protocol (e.g., GOSSIP1($p$) in [14]). When a mobile node first receives an RREQ packet, it will either broadcast the RREQ packet to its neighbors with probability $p$ or discard it with probability $1 - p$. Therefore, an RREQ packet will be broadcasted over more than one hop with probability $1 - (1 - p)^n$, where $n$ is the number of neighbors. In our simulation model, we use the modified GOSSIP1$(p, k)$. In GOSSIP1$(p, k)$, when a mobile node first receives an RREQ packet, with probability 1 it will broadcast the RREQ packet to its neighbors for the first $k$ hops. However, after $k$ hops from the source node $S$, GOSSIP1$(p, k)$ works exactly the same way as GOSSIP1$(p)$. Other variations of the gossip-based routing protocols have been proposed recently (e.g., [20] and [21]). Performance comparisons between LB-AODV and these protocols are subject of future work.

Consider the LB-AODV routing protocol. When a source node sends an RREQ message, $m$ out of $n$ neighboring nodes will broadcast the RREQ packet to its neighbors, while the other neighboring nodes will discard the packet. Therefore, an RREQ packet can be broadcasted over more than one hop with probability $p = 1$ if the group number is chosen correctly. Since the LB-AODV routing protocol regulates the number $m$ dynamically, it can control the number of RREQ packets being broadcasted without degrading the level of network connectivity.

## III. Simulation Model and Evaluations

In this section, we compare the performance between our proposed LB-AODV, the original AODV [9], and the gossip-based routing [GOSSIP1$(p, 1)$] [14] protocols.

### A. Simulation Model

The network simulator $(ns2)$ [19] is used for the implementation of LB-AODV and GOSSIP1 routing protocols. The physical radio characteristics of each mobile node's radio interface are chosen to approximate the Lucent WaveLAN [22] operating as a shared-medium radio with a nominal bit rate of 2 Mb/s and a nominal radio range of 250 m. For the medium access control layer, the IEEE 802.11 distributed coordination function (DCF) [2] is used. The propagation model combines both a free space propagation model and a two-ray ground reflection model. We use the same configuration parameters as those of *ns2* version b8a.

Constant bit rate (CBR) traffic sources are used with different packet generation rates. The data packet size is 512 bytes. The size of the network is $1500 \times 300$ $(\mathrm{m}^2)$, and the number of mobile nodes is 50 in the simulations which results are presented in Figs. 4–7. On the other hand, a $1500 \times 600$ $(\mathrm{m}^2)$ topology is used with 100 mobile nodes in the simulations which results are shown in Fig. 8. A $1000 \times 1000$ $(\mathrm{m}^2)$ topology is also used with various number of mobile nodes in simulations that yielded the results in Fig. 9. Table I provides a summary of the simulation parameters. One stationary gateway node is located in the middle of the grid [i.e., coordinate (750, 150)] for the first three simulation scenarios. A *random waypoint model* [23] is used for the mobility model. Each node moves at a speed that is uniformly distributed from 0 to 20 m/s. Each simulation run takes 900 simulated s. The results presented are mean values of at least ten simulation runs. For fair comparisons, all three routing protocols use the same set of mobility and traffic scenarios.

For comparisons with gossip-based routing, since only $T$ mobile nodes can relay packets generated by each group in LB-AODV, we choose the gossip probability $p$ to be equal to $T/M$. Thus, after $k$ hops from the source node, when a neighboring node first receives an RREQ packet, it will either broadcast the RREQ packet to its neighbors with probability $T/M$, or discard it with probability $1 - T/M$. Table II provides a summary of the values of $T$, $G$, and $p$ by varying the number of source nodes $S$. The value of $k$ is chosen as 1 because the average path length is about 2.5.

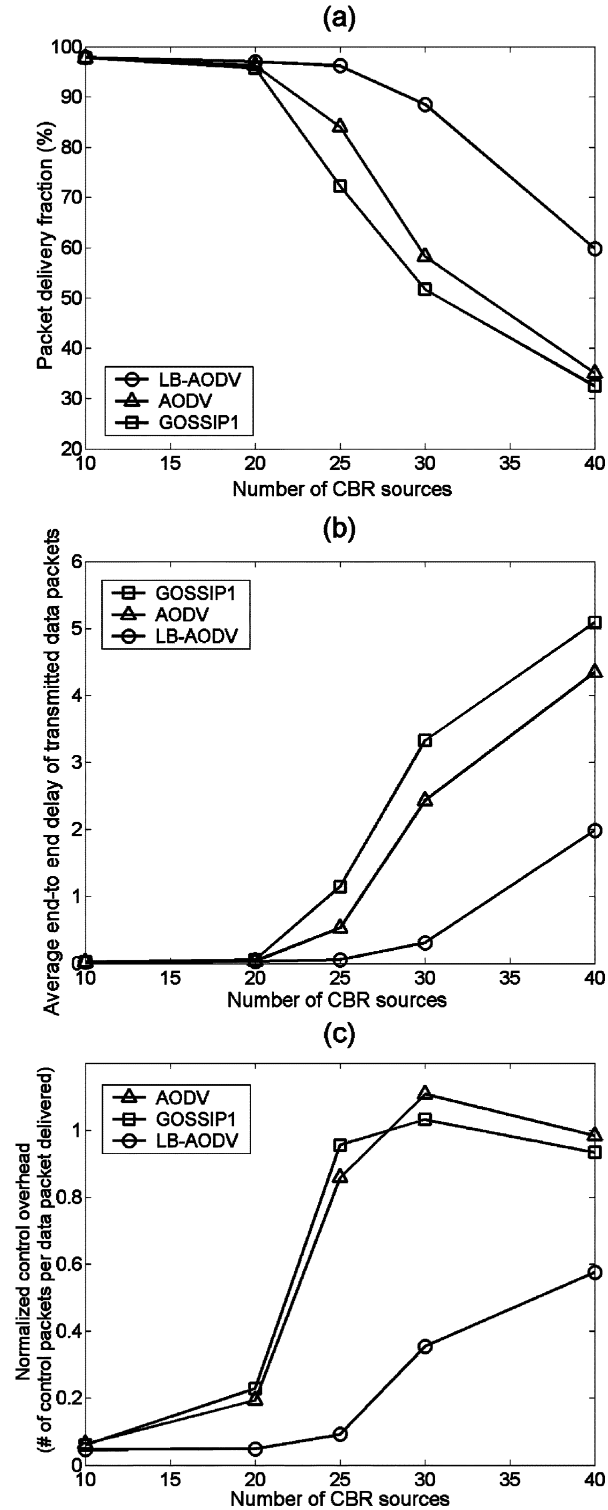



Fig. 4. Performance comparisons between AODV, LB-AODV, and GOSSIP1 routing protocols with varying number of CBR sources (pause time $= 500$ s).

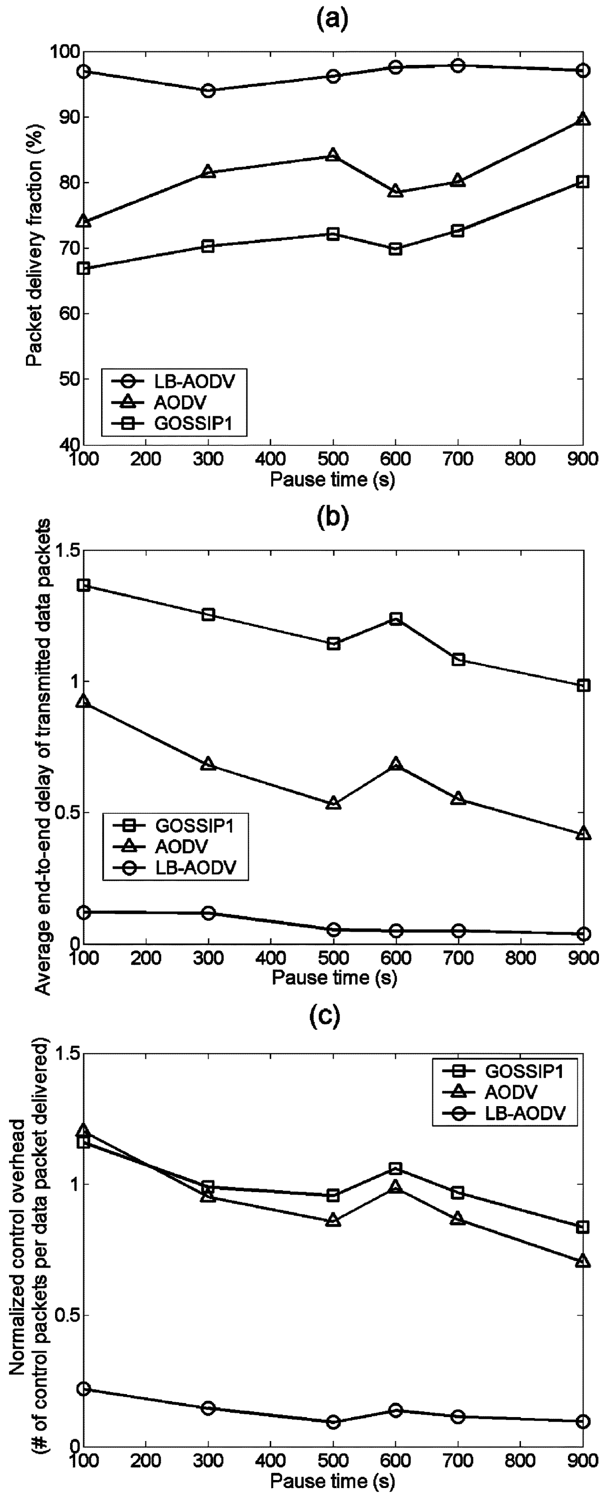



Fig. 5. Performance comparisons between AODV, LB-AODV, and GOSSIP1 routing protocols over a range of pause time (number of CBR sources = 25).

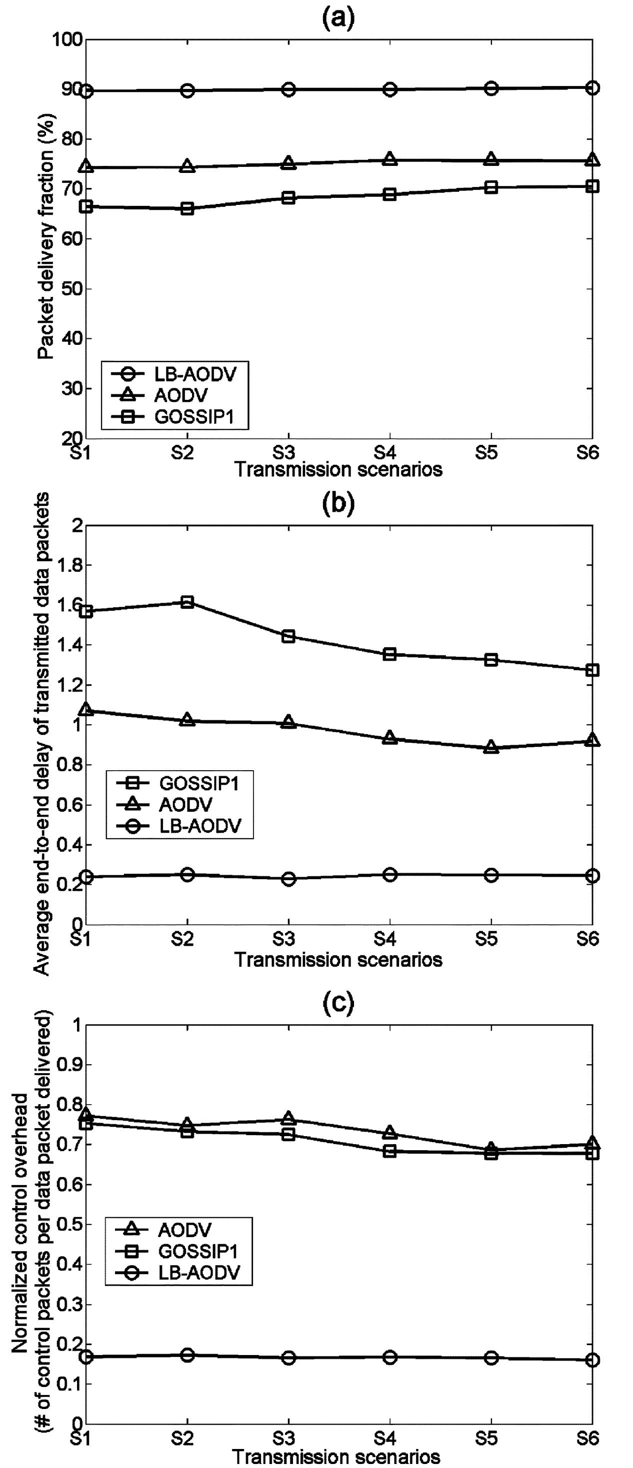



Fig. 6. Performance comparisons between AODV, LB-AODV, and GOSSIP1 routing protocols for variable source rate scenarios shown in Table III (pause time = 500 s).

### B. Performance Metrics

The following performance metrics are used for comparisons. The *packet delivery fraction* is defined as the measured ratio of the number of data packets delivered to the destinations to the number of packets generated by all traffic sources. The *average end-to-end delay of transferred data packets* includes all possible delays caused by buffering during route discovery, queueing at the interface-queue, retransmission delays at the medium access control layer, and propagation and transfer times. The *normalized control overhead* is defined as the number of both routing and update (in LB-AODV) packets transmitted per data packet delivered at the destination. Note that each time a packet is forwarded is counted as one packet transmission.

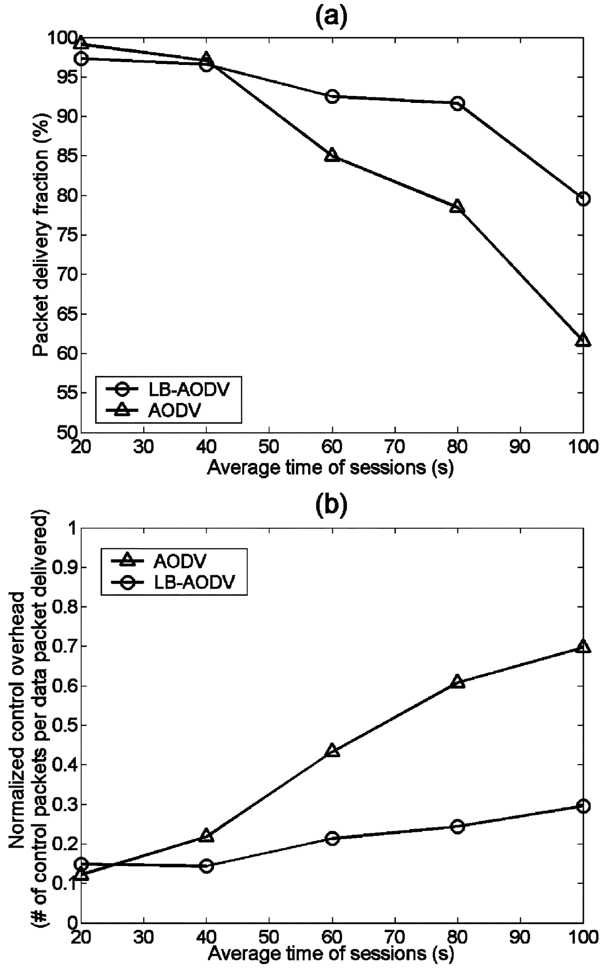



Fig. 7. Performance comparisons between AODV and LB-AODV routing protocols with variable numbers of source nodes (average intersession time = 60 s).

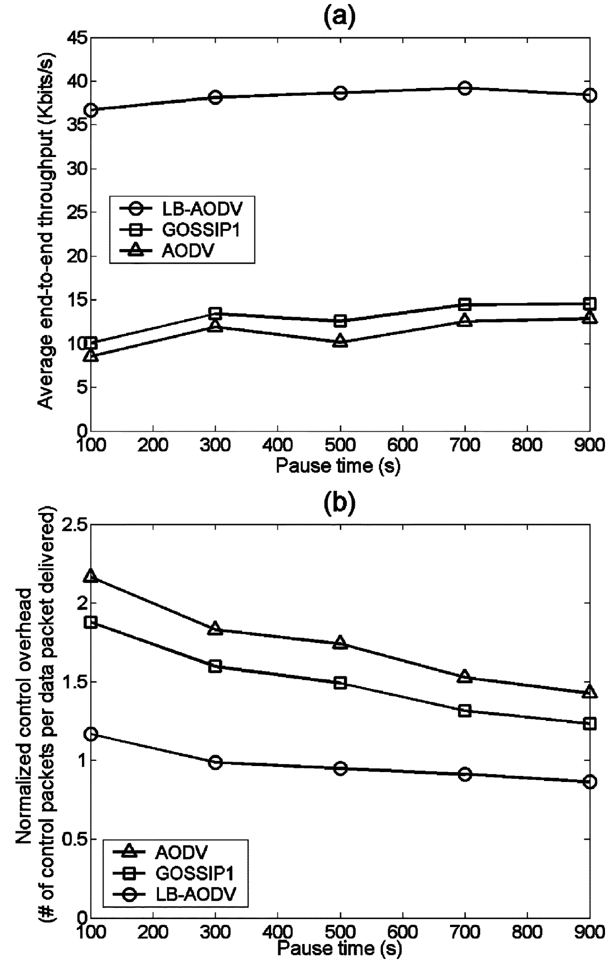



Fig. 8. Performance comparisons between AODV, LB-AODV, and GOSSIP1 routing protocols in a two-gateway scenario (number of CBR sources = 40).

### C. Performance Comparisons

*Scenario 1: Single Gateway, Multiple Source Nodes With Same Packet Generation Rate:* Fig. 4 shows the performance of the network with different number of CBR sources. When the number of sources is less than 20, all three routing protocols provide a high packet delivery fraction, small end-to-end delay, and normalized control overhead. Results in Fig. 4(a) indicate that LB-AODV improves the packet delivery fraction by 15% over the other schemes when the number of sources increases to 25. As traffic further increases, the improvement is increased radically. This implies that when traffic load is high (i.e., more than 25 sources in this scenario), most of the routes toward the gateway are congested by a lot of control and data packets. Therefore, contention and collision between neighbors increase exponentially and, thus, the AODV and GOSSIP1 routing schemes become less efficient. Results in Fig. 4(b) indicate that within a given end-to-end delay constraint, LB-AODV can support more traffic when compared with the other protocols. Fig. 4(c) shows that LB-AODV has a much lower normalized control overhead when compared with AODV and GOSSIP1 routing protocols.

Fig. 5 shows the overall performance by varying pause time (i.e., mobility). The number of source nodes is equal to 25. These results indicate that in a slightly congested network (with 25 source nodes) LB-AODV maintains a better performance over different mobility rates when compared with AODV and GOSSIP1.

*Scenario 2: Single Gateway, Multiple Source Nodes With Different Packet Generation Rates:* Our proposed load-balancing mechanism distributes the number of source nodes evenly among different groups. Therefore, it cannot balance the average packet transmission rates of each group if each source node has a different packet generation rate. In this simulation, we investigate the effects of source nodes with different packet generation rates on the performance of LB-AODV routing protocol. Table III provides six cases where source nodes with different packet generation rates of three, six, and nine packets/s are mixed. The pause time is equal to 500 s in this scenario. Note that for fair comparisons the average packet generation rate in each scenario is equal to 120 Kb/s.

Fig. 6 compares the performance of AODV, GOSSIP1, and LB-AODV for the six scenarios shown in Table III. As shown in Fig. 6(a)–(c), the performance of LB-AODV is almost constant among different scenarios. Moreover, in all the scenarios considered, LB-AODV consistently and significantly outperforms AODV and GOSSIP1 routing protocols. Results in Fig. 6 confirm that the performance gain of LB-AODV over AODV and

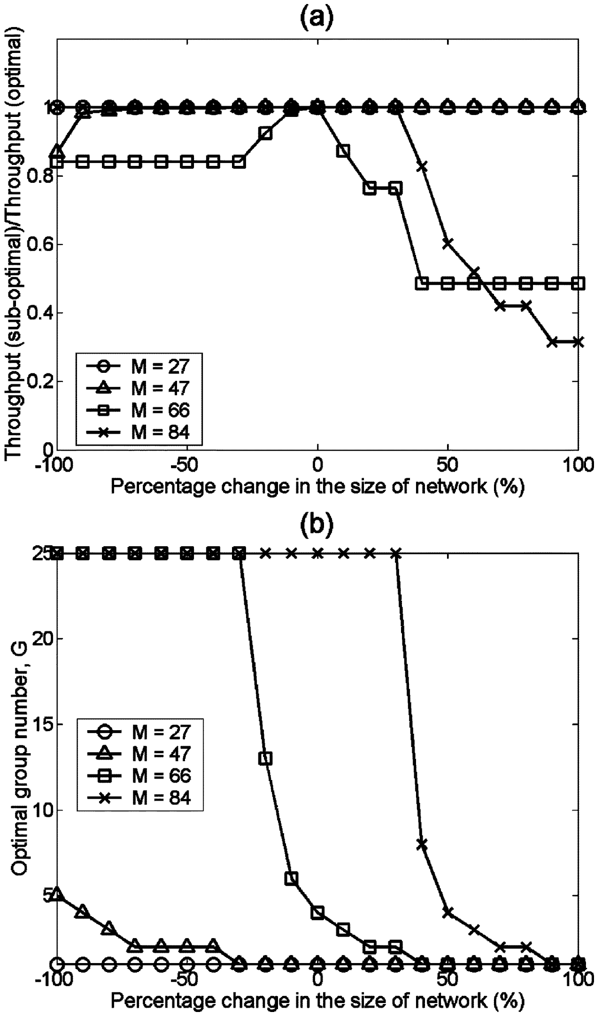



Fig. 9. Sensitivity analysis of the estimated size of network (number of CBR sources $= 25$, pause time $= 500$ s).

TABLE I
SIMULATION PARAMETERS

| | | | |
|---|---|---|---|
| transmission range | 250 m | topology size | 1500×300 $m^2$<br>1500×600 $m^2$<br>1000×1000 $m^2$ |
| bandwidth of radio interface | 2 Mb/sec | traffic type | CBR |
| simulation time | 900 sec | packet generation rate | 3, 4, 6, 8, 9 packets/sec |
| no. of nodes | 27, 47, 50, 66, 84, 100 | packet size | 512 Bytes |
| no. of source nodes | 10, 20, 25, 30, 40 | pause time (sec) | 100, 300, 500, 600, 700, 900 |

GOSSIP1 is caused mainly by the reduction of the number of control packet transmissions. Thus, LB-AODV is also efficient in mobile ad hoc wireless access networks that are composed of source nodes with different packet generation rates.

*Scenario 3: Single Gateway and Variable Number of Source Nodes:* When the number of source node changes, the gateway has to update the value of the balance index and broadcast an advertisement message to all the nodes to update the state information. Therefore, it is expected that as the number of source node changes frequently, a considerable number of control packets are propagated in the network for state information update. In this set of simulations, we investigate the effects of broadcasting of control packet in LB-AODV. The pause time is set to 500 s, and the maximum number of source nodes is 25. The packet generation rate for each source node is 8 packets/s. By changing the average session time between communication pairs, we can obtain the results for different traffic densities. Both the average session time and the average intersession time are assumed to follow the exponential distribution. As described in Section II-C, when the gateway has not received data packets corresponding to a particular entry for a certain period of time (10 s in this case), the corresponding routing entry and its group number will be deleted.

TABLE II
SIMULATION VARIABLES

| *variables* / *no. of sources, S* | *no. of possible relay nodes, T* | *no. of groups, G* | *gossip probability, p* |
|---|---|---|---|
| 10 | 41 | 10 | 41/50 |
| 20 | 31 | 20 | 31/50 |
| 25 | 30 | 5 | 30/50 |
| 30 | 30 | 3 | 30/50 |
| 40 | 30 | 2 | 30/50 |

TABLE III
TRANSMISSION SCENARIOS

| | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|
| *no. of source nodes with 3 packets/sec* | 0 | 2 | 4 | 6 | 8 | 10 |
| *no. of source nodes with 6 packets/sec* | 20 | 16 | 12 | 8 | 4 | 0 |
| *no. of source nodes with 9 packets/sec* | 0 | 2 | 4 | 6 | 8 | 10 |

Fig. 7 shows the performance by varying the average session time. The average intersession time is set to 60 s. Fig. 7(a) and (b) shows that when the average session time is less than 40 s, AODV works slightly better than LB-AODV with a lower control overhead. On the other hand, LB-AODV outperforms AODV as the average session time increases. We observe that LB-AODV has a lower normalized control overhead as the average session time increases. Note that as the traffic density increases, with AODV, the network becomes congested with routing and data packets. On the other hand, the grouping mechanism in LB-AODV controls the amount of routing control overhead. It remains efficient even when the number of source nodes changes.

*Scenario 4: Two Gateways:* This experiment relates to the study of scalability with two gateways. We determine the variation of the throughput (i.e., the total amount of bytes received without errors by the destination per second) and the normalized control overhead by increasing the network size to a $1500 \times 600$ $m^2$ topology and changing the number of mobile nodes to

100. We consider 40 CBR sources, each with a packet generation rate of 4 packets/s. Since the number of common nodes $M - S = 60$ exceeds the optimal number $R = 50$ for this topology, we choose the maximum number of source nodes, 40, as the total number of groups $G$. Therefore, each source node belongs to a different group (refer to Section II-D). The simulation time is 900 s. All the other simulation parameters remain the same.

In this scenario, two gateways, $G1$ and $G2$, are located in the coordinates (750, 150) and (750, 450), respectively. Since each gateway can monitor the number of source nodes being served, each gateway communicates with 20 source nodes at a maximum. In LB-AODV, the total number of mobile nodes $T$ that relay packets generated by each group is 61 (i.e., $M - S + (S/G) = 100 - 40 + (40/40) = 61$). The *state information* should be a $(G + 2)$—tuple in the form of $\langle gateway\ number,\ group\ number, f_1, f_2, \ldots, f_G \rangle$ in this scenario.

Fig. 8(a) shows the throughput as a function of pause time in the network. Since LB-AODV can divide only source nodes into different groups, the increase of control overhead is unavoidable as the number of mobile nodes increases. However, due to the fact that the grouping mechanism can reduce the amount of routing control overhead [see Fig. 8(b)] and distribute the number of source nodes between two gateways, the throughput of LB-AODV is approximately three times higher than that of AODV and GOSSIP1 routing protocols. These results show that LB-AODV is still efficient in scenarios with two gateways and a large network size. Note that GOSSIP1 shows a better performance than AODV in this scenario. This is because in large and dense networks gossip-based routing protocols are effective in improving the efficiency by reducing the transmissions of routing control packets [14]. Further performance improvements for LB-AODV may be possible by refining the group assignment algorithm to take into account of the number and location of gateways. This is a subject for further research.

*Scenario 5: Sensitivity Analysis:* Recall that the optimal group number $G$ is a function of $M$, $S$, and $R$ [see (2), in Section II-D]. Although the parameters $M$ and $S$ can be monitored by the gateway, the value of $R$ may not always be estimated correctly. If that is the case, the resulting number $G$ may not indeed be optimal in terms of node density. We are interested in determining the percentage change of the throughput as a function of the variations of the size of network $Z$. The procedures for the sensitivity analysis consist of the following steps.

Step 1) Given the actual size of the network $Z$, we first determine the optimal value of $R$.

Step 2) Given the values $R$, $M$, and $S$, we determine the optimal group number $G$ based on (2).

Step 3) Given the values $R$, $M$, $S$, and $G$, the expected throughput can be obtained via the *ns2* simulation. We denote the value as *throughput (optimal)*.

Step 4) Let $Z'$ denote the estimated size of network and $\Delta_Z$ denote the percentage change of the size of network. These parameters are related by the following equation:

$$\mathrm{Z}' = (1 + \Delta_{\mathrm{Z}})\mathrm{Z} \tag{3}$$

Based on the estimated size of network $Z'$, the suboptimal value of $R'$ is determined. Similarly, given the values $R'$, $M$, and $S$, the suboptimal group size $G'$ can be calculated based on (2). The suboptimal expected throughput, denoted as *throughput (suboptimal)*, is obtained via the *ns2* simulation.

Step 5) The change of the throughput with respect to the variation of the size of network is characterized by the *throughput ratio*, which is defined as: *throughput (suboptimal)/throughput (optimal)*.

Fig. 9(a) shows the throughput ratio versus the percentage change of the size of network. We assume the actual topology of the network to be $1000 \times 1000\ \mathrm{m}^2$. The optimal value of $R$ is 47. There is one stationary gateway located in the coordinate (500, 500). When the size of network is underestimated by 100%, the suboptimal value of $R'$ is 27. On the other hand, the suboptimal value of $R'$ is 84 when the size of network is overestimated by $100\%$. To study the effect of the number of nodes to the estimated size of network, we vary the number of nodes from 27 to 84. Note that the number 27 and 84 are the suboptimal values of $R$ when the estimation is deviated by $-100\%$ and $+100\%$, respectively. Fig. 9(b) shows that the number of groups $G$ based on the given values of $M$, $S$, and $R$.

When the number of node $M$ is less than or equal to 47, the throughput ratio is not sensitive to the estimated size of network. On the other hand, as the number of nodes $M$ increases, the throughput ratio is sensitive to both underestimation and overestimation of the size of network. An overestimation of the size of network gives a higher throughput ratio than an underestimation of the same percentage. These results imply that if there is uncertainty in estimating the size of network, it may be better to underestimate its value in order to reduce the throughput ratio difference.

## IV. Conclusion

With flooding-based on-demand route discovery in mobile ad hoc wireless access networks, many routing messages (i.e., RREQ) are propagated unnecessarily. Moreover, the redundancy of routing information (i.e., RREP and RREQ) processed by the gateway is high in the mobile ad hoc wireless access network. To reduce the overhead of routing messages, we have proposed an extension of the ad hoc on-demand routing protocol by incorporating the concept of load-balancing in this paper. Our proposed LB-AODV protocol is simple and well-suited for the mobile ad hoc wireless access network environment.

We have compared the performance of our proposed LB-AODV protocol with both the original AODV and gossip-based routing protocols in different mobility and traffic scenarios. Simulation results show that LB-AODV delivers more data packets to the gateway and decreases the end-to-end delay of packets delivered by reducing the transmissions of routing control messages by 50% or more. In scenarios with traffic congestion, LB-AODV significantly outperforms AODV and GOSSIP1 routing protocols. We have compared the performance of the protocols in a scenario with a larger number of mobile nodes accessing two gateways. LB-AODV provides significant advantages over AODV and GOSSIP1 in terms of throughput and routing overhead even in a large network with two gateways. Although we have presented the details of LB-AODV based on the AODV routing protocol, the load-balancing concept developed in this paper can generally be applied to other on-demand routing schemes.

To facilitate practical implementation of our proposal, we are investigating techniques that provide good estimations of network size and topology in a dynamic MANET. We are also seeking further improvements of our group assignment mechanism, especially for large networks with multiple gateways. Furthermore, we are considering how our load-balancing concept can be incorporated in other on-demand routing protocols with different routing metrics (e.g., the least load or least power route).

## REFERENCES


[1] S. Corson and J. Macker, "Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations," IETF RFC 2501, Jan. 1999.

[2] *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, ISO/IEC 8802-11; ANSI/IEEE Standard 802.11, Aug. 1999.

[3] Bluetooth Core Specification, v1.2, Nov. 2003.

[4] C. E. Perkins, J. T. Malinen, R. Wakikawa, A. Nilsson, and A. J. Tuominen, "Internet connectivity for mobile ad hoc networks," *J. Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 465–482, Aug. 2002.

[5] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An analysis of the optimum node density for ad hoc mobile networks," in *Proc. IEEE ICC*, Helsinki, Finland, June 2001, pp. 857–861.

[6] M. Sanchez, P. Manzoni, and Z. J. Haas, "Determination of critical transmission range in ad-hoc networks," in *Proc. Multiaccess Mobility and Teletraffic Wireless Communications 1999 Workshop (MMT)*, Venice, Italy, Oct. 1999, pp. 293–304.

[7] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," in *Proc. IEEE Nat. Telecommunications Conf.*, 1978, pp. 431–435.

[8] S. T. Sheu and J. Chen, "A novel delay-oriented shortest path routing protocol for mobile ad-hoc networks," in *Proc. IEEE ICC*, Helsinki, Finland, June 2001, pp. 1930–1934.

[9] C. E. Perkins, E. Belding-Royer, and S. R. Das, "Ad-hoc on-demand distance vector (AODV) routing," IETF RFC 3561, July 2003.

[10] D. B. Johnson, D. A. Maltz, and Y. C. Hu, "The dynamic source routing protocol for mobile ad hoc networks (DSR)," IETF Internet Draft (work in progress), July 2004.

[11] J. Li, C. Blake, S. J. Douglas, D. Couto, H. Lee, and R. Morris, "Capacity of ad-hoc wireless network," in *Proc. ACM MOBICOM*, Rome, Italy, Sept. 2001, pp. 61–69.

[12] S. Roy and J. J. Garcia-Luna-Aceves, "Node-centric hybrid routing for ad-hoc wireless extensions of the Internet," in *Proc. IEEE GLOBECOM*, Taipei, Taiwan, Nov. 2002, pp. 183–187.

[13] S. Y. Ni, Y. C. Tseng, Y. S. Chan, and J. P. Sheu, "The broadcast storm problem in a mobile ad-hoc network," in *Proc. ACM MOBICOM*, Seattle, WA, Aug. 1999, pp. 151–162.

[14] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad-hoc routing," in *Proc. IEEE INFOCOM*, New York, June 2002, pp. 1707–1716.

[15] Y. Yi, M. Gerla, and T.-J. Kwon, "The selective intermediate nodes scheme for ad-hoc on-demand routing protocols," in *Proc. IEEE ICC*, New York, Apr./May 2002, pp. 3191–3196.

[16] J.-H. Song, V. Wong, and V. Leung, "Efficient on-demand routing for mobile ad-hoc wireless access networks," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Dec. 2003, pp. 558–563.

[17] P. Hsiao, A. Hwang, H. Kung, and D. Vlah, "Load-balancing routing for wireless access networks," in *Proc. IEEE INFOCOM*, Anchorage, AK, Apr. 2001, pp. 986–995.

[18] M. R. Pearlman and Z. J. Haas, "Determining the optimal configuration for the zone routing protocol," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1395–1414, Aug. 1999.

[19] The network simulator—NS-2 Notes and documentation and source code. [Online]. Available: http://www.isi.edu/nsnam/ns/

[20] J. Luo, P. T. Eugster, and J.-P. Hubaux, "Route driven gossip: Probabilistic reliable multicast in ad-hoc networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar./Apr. 2003, pp. 2229–2239.

[21] Y. Sasson, D. Cavin, and A. Schiper, "Probabilistic broadcast for flooding in wireless mobile ad-hoc networks," in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 2003, pp. 1124–1130.

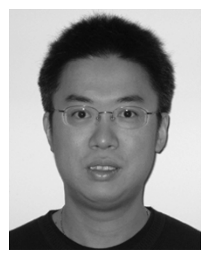
[22] B. Tech, "Development of WaveLAN, an ISM band wireless LAN," *AT&T Tech. J.*, pp. 27–37, July/Aug. 1993.

[23] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad-hoc network routing protocols," in *Proc. ACM MOBICOM*, Dallas, TX, Oct. 1998, pp. 85–97.



**Joo-Han Song** (S'99) received the M.S. degree in electrical engineering from the Hongik University, Seoul, Korea, in 2001. Currently, he is working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of British Columbia (UBC), Vancouver, BC, Canada, where he is a Member of the Communications Group.

His research interests include routing and security for mobile ad hoc networks, and the performance evaluation and modeling of wireless networks.


**Vincent W. S. Wong** (S'94–M'00) received the B.Sc. (with distinction) degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000, all in electrical engineering.

From 2000 to 2001, he was a Systems Engineer at PMC-Sierra, Inc., Burnaby, BC. Since 2002, he has been with the Department of Electrical and Computer Engineering, UBC, where he is currently an Assistant Professor. His research interests are in wireless communications and networking.

Dr. Wong received the Natural Science and Engineering Research Council (NSERC) postgraduate scholarship and the Fessenden Postgraduate Scholarship from Communications Research Centre, Industry Canada, during his graduate studies.



**Victor C. M. Leung** (S'75–M'89–SM'97–F'03) received the B.A.Sc. (Honors) degree in electrical engineering from the University of British Columbia (UBC), Vancouver, BC, Canada, in 1977 and the Ph.D. degree in electrical engineering from UBC in 1981 on a Natural Sciences and Engineering Research Council Postgraduate Scholarship.

From 1981 to 1987, he was Senior Member of Technical Staff at Microtel Pacific Research Ltd. (later renamed MPR Teltech Ltd.),Burnaby, BC, specializing in the planning, design, and analysis of satellite communication systems. He also held a part-time position as Visiting Assistant Professor at Simon Fraser University, Burnaby, BC, in 1986 and 1987. In 1988, he was a Lecturer in the Department of Electronics, Chinese University of Hong Kong. He joined the Department of Electrical Engineering at UBC. in 1989, where he is a Professor, holder of the TELUS Mobility Industrial Research Chair in Advanced Telecommunications Engineering, and a Member of the Institute for Computing, Information and Cognitive Systems. He was a Project Leader and a Member of the Board of Directors in the Canadian Institute for Telecommunications Research, a Network of Centres of Excellence funded by the Canadian Government. His research interests are in the areas of architectural and protocol design and performance analysis for computer and telecommunication networks, with applications in satellite, mobile, personal communications, and high-speed networks.

Dr. Leung was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science in 1977. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is a Voting Member of the Association for Computing Machinery (ACM).