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

Joo-Han Song, Vincent W. S. Wong and Victor C. M. Leung

Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, BC, Canada V6T 1Z4
e-mail: {jooahans, vincentw, vleung}@ece.ubc.ca

Abstract— In this paper, we consider a mobile ad-hoc wireless access network in which mobile nodes can access the Internet via a stationary gateway node or access point. Mobile nodes that are outside the transmission range of the gateway can continue to communicate with the gateway via a multi-hop connection with their neighboring nodes. The Ad-hoc On-demand Distance Vector (AODV) routing protocol is extended by incorporating the concept of load-balancing (LB). We call this the LB-AODV routing protocol. Simulation results show that in a congested network environment, our proposed LB-AODV has a higher packet delivery fraction, a lower end-to-end delay and control overhead when compared with both AODV and gossip-based routing protocols.

Keywords - *mobile ad hoc wireless access network; load-balancing on-demand routing protocol*

I. INTRODUCTION

A Mobile Ad-hoc NETWORK (MANET) consists of a set of wireless mobile nodes communicating to each other without any centralized control or fixed network infrastructure. MANETs have been evolving to serve a growing number of applications that rely on fast deployable multi-hop wireless infrastructures.

Over the past few years, a number of ad-hoc routing protocols have been proposed in the literature. These protocols can be divided into two categories [1]: proactive and on-demand. Proactive routing protocols make use of periodic routing updates while on-demand routing protocols only maintain the routes that are currently in use. For on-demand routing protocols, route discovery and maintenance are performed via the exchange of Route Request (RREQ), Route Reply (RREP) and Route Error (RERR) control packets. Since on-demand routing protocols react quickly to topology changes, they perform better than proactive routing protocols for normal network scenarios in MANETs.

In this paper, we consider a mobile network environment in which mobile hosts can access the Internet via a gateway or access point. We refer to this as a *mobile ad-hoc wireless access network*. Mobile hosts that are near the gateway can communicate directly with the gateway via a single hop connection. However, mobile hosts that are outside the transmission range of the gateway have to use multi-hop connections that rely on the neighboring mobile nodes to relay

Figure 1. Mobile ad-hoc wireless access network.

their packets (see Figure 1).

In general, the metrics used to evaluate the performance of ad-hoc routing protocols in MANETs include packet delivery fraction, end-to-end delay, and routing control overhead. These performance metrics depend on a number of factors including the total number of mobile nodes in the network, the number of source nodes (i.e., those which have data packets to send), the traffic load, the movement pattern of the mobile nodes, the pause time between moves, and the transmission range.

Consider the effects of the total number of mobile nodes in a MANET over a given coverage area. Results in [2] showed that when the number of nodes is small, the network may not be fully connected. Some nodes may not be able to send packets to certain destinations. However when the number of nodes and the traffic load increase, results in [3] show that contention and packet collision between neighboring nodes may increase exponentially. In addition, the routing control overhead also increases with the number of nodes. When a source node needs to send the RREQ message to all of its neighboring nodes during route discovery, the total number of RREQ messages is proportional to the total number of mobile nodes in the network. When the network is congested, routing control packets (e.g., RREP) may be dropped. The source node may initiate another route discovery process, which further increases the amount of control traffic in the network [4][5].

In order to maintain a high packet delivery fraction and a low end-to-end delay for packet transmission, it is important to reduce the amount of routing control information in the network. In this paper, we propose the use of load-balancing to limit the amount of routing control packets in mobile ad-hoc wireless access networks. In our proposed load-balancing scheme, route selection is regulated by a distributed grouping

mechanism, which distributes data traffic and reduces unnecessary routing traffic. The parameter for grouping can be continuously updated by the gateway.

The rest of this paper is organized as follows. Our proposed extension of the Ad-hoc On-demand Distance Vector routing protocol with Load-Balancing (LB-AODV) is described in Section II. The performance comparisons between our proposed LB-AODV, the original AODV [6], and the gossip-based routing [7] 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 operation of LB-AODV and the route selection process. This is followed by a discussion on the selection of the total number of groups.

The following terminologies will be used:

- *Source node*: a mobile node with data packets to send towards the gateway.
- *Common node*: a mobile node that is neither a source nor destination node, and does not belong to any group.
- *Active node*: a mobile node that has valid route(s) to the gateway and is currently being used to forward packets towards the gateway.

A. Load-Balancing Mechanism

Since on-demand routing protocols are based on flooding, many routing messages are propagated unnecessarily. In order to maintain a high performance in data packet transmission, it is important to reduce the amount of routing messages. The proposed load-balancing mechanism reduces the number of unnecessary retransmissions of routing messages and prevents heavy network congestion by separating source nodes into different *groups* and allowing source nodes to relay packets generated only by their own group members.

Our grouping mechanism partitions all mobile nodes into several logical divisions such as *A*, *B*, *C*, *D*, and *E* as shown in the example in Figure 2. All common nodes, which are assigned to division *E*, are allowed to relay packets from any groups towards the gateway. On the other hand, the source nodes of any groups are not allowed to relay packets from other groups. 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 the division *E*. Therefore, the maximum number of retransmissions of RREQ message, which is generated by any group, is reduced by the number of source nodes of other groups at the route discovery process. The decision of the number of groups is important for the operation of LB-AODV routing protocol. It is discussed in detail in Section II-C.

A source node that is located in the shortest path between other source nodes and the gateway would be easily congested with routing and data messages of its own and others. The traffic congestion may lead to a higher packet dropping rate and faster battery power depletion. By dividing source nodes into several groups, the packet relaying responsibility and the traffic load can be balanced among different groups, and thus

Figure 2. An example of logical partitioning of mobile nodes.

the maximum level of congestion is reduced at the sources.

The proposed load-balancing mechanism aims at maximizing the balance index, *B*, which is defined as [8]:

$$B = \left(\sum_{i=1}^G f_i \right)^2 / \left(G \sum_{i=1}^G f_i^2 \right) \quad (1)$$

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

For simplicity and robustness, we are considering a load-balancing algorithm that distributes the number of source nodes evenly among different groups. For future work, we plan to extend eqn. (1) by including the average packet transmission rate in each group.

B. Load-Balancing Routing Process

Using the load-balancing route discovery process, we can minimize the variance of the total number of source nodes between groups dynamically. The flow charts for the route selection process are shown in Figure 3. A group number is assigned only to each source node that initiates the route discovery process.

When a source node has data to send but does not know a route to the gateway, it initiates the route discovery process by broadcasting the RREQ message to its neighboring nodes (see Figure 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. Note that the state information in each active node is synchronized periodically with that in the gateway as described below. 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 RREP message to the source node. This RREP message includes the 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 Figure 3(b). Since an active intermediate node can assign different groups to different source nodes, it needs to maintain different route entries to the gateway for different groups it is currently serving.

Figure 3. Processes in mobile nodes and gateway.

Similarly, when the gateway node receives a 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 a RREP message to the source node. The flow chart for the operations of the gateway node is shown in Figure 3(c). When the source node receives a RREP message, it will begin sending data packets to the gateway immediately via the node from which it received the RREP message.

The gateway monitors the total number of source nodes from each group periodically. Whenever the change in the total number of source nodes from any group has exceeded a certain threshold, the value of the balance index needs to be updated. The gateway then sends an advertisement message to all the active nodes to update the state information.

We now describe the procedures for route maintenance. When a source node detects a link breakage via the RERR message, it will re-initiate the route discovery by sending a RREQ message with its group number towards the gateway. Those intermediate nodes that do not belong to this particular

group will simply drop the RREQ message. When 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 a RREP 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. In the worst case, if the source node has not received any RREP message after a certain period of time, it will re-initiate the route discovery process again by sending another RREQ message without the group number.

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

C. Discussions

As the number of mobile nodes increases, the contention and packet collision between neighboring nodes increase exponentially [3]. In addition, the amount of routing control overhead also increases. On the other hand, when the number of nodes is small, the network may not be fully connected. Therefore, the determination of the number of groups is critical for the efficiency of LB-AODV routing protocol. The number of groups is chosen as a trade-off between the network connectivity and the amount of routing control overhead. To decide on the number of groups, the gateway has to know the number of source nodes, the number of mobile nodes, and the network topology. The estimation of these parameters is beyond the scope of this paper. We would like to point out that the gateway can estimate the number of source nodes by monitoring the source address field in the packet header. The gateway can estimate the number of mobile nodes and the size of the network by the number of neighbors around the gateway and averaging the hop count seen in the RREQ packet.

Results in [2] have shown that for normal network scenarios in MANETs, the best performance can be achieved when the average number of neighbors is around 7. In this paper, we define the optimal number of mobile nodes R in a given topology as the number of mobile nodes that results in the average number of neighbors being around 7. Since the gateway can estimate the size of the network, it can calculate the optimal number of mobile nodes R in the estimated topology.

We aim to minimize the difference between the optimal number of mobile nodes R and the total number of mobile nodes T , which can relay packets generated by each group. The rationale behind this operation is the fact that the optimal number R gives the best performance without decreasing the network connectivity in a given topology. Given the number of source nodes S , the number of mobile nodes M , and the total number of groups G , the total number of common nodes comes to $M - S$, and the number of source nodes belonging 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 by

minimizing the absolute difference between T and R :

$$G = \arg \min_g |T - R|$$

$$= \arg \min_g |M - S + (S/g) - R| \quad (2)$$

where g is an integer and $1 \leq g \leq S$.

Consider the following example, in a given network topology with a size of $1500 \times 300 \text{ m}^2$, the network simulation (ns2) “setdest” [9] can be used to calculate R such that the average number of neighbors is around 7. Based on this calculation, the value of 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 based on eqn. (2), the number of groups G is equal to 5.

III. SIMULATION MODEL AND EVALUATIONS

In this section, we compare the performance between our proposed LB-AODV, the original AODV [6], and the gossip-based routing (GOSSIP) [7] routing protocols.

We begin by reviewing the gossip-based routing protocol. Gossip-based routing aims to reduce the routing control overhead by selectively discarding some control packets. A source node sends the RREQ packet with probability 1. When a node first receives a RREQ packet, with probability p it broadcasts the RREQ packet to its neighbors, and with probability $1 - p$ it discards this RREQ packet. The authors in [7] recommended the use of probability 1 for the first k hops before continuing with probability p for subsequent hops. This version of gossip-based routing protocol is called *GOSSIP1(p, k)*. We use this gossip-based routing protocol for performance comparisons.

A. Simulation Model

A stationary gateway node is located in the middle of the grid. A *random waypoint model* [10] is used for the mobility model. The speed of the mobile is uniformly distributed from 0 to 20 m/s. The simulation time is 900 seconds. The results are presented on the basis of at least 5 simulation runs. For fair comparisons, several sets of identical mobility and traffic scenarios are applied to all routing protocols at the same time.

The Network Simulator (ns2) [9] is used for performance evaluations. The physical radio of each mobile node’s radio interface is chosen to approximate the Lucent WaveLAN [11], operating as a shared-media radio with a nominal bit rate of 2 Mbits/s and the radio range of 250 m. For the MAC layer, the IEEE 802.11 Distributed Coordination Function (DCF) [12] is used. The propagation model combines both a free space and a two-ray ground reflection model. We use the same configuration parameters as those of ns2 version b8a.

The size of the network is $1500 \times 300 \text{ m}^2$. The number of mobile nodes M is 50. Each source node S generates constant bit rate (CBR) traffic with packet generation rate of 4 packets per second. Table 1 provides a summary of the simulation parameters.

Given the network size, the ns2 “setdest” [9] is used to calculate R such that the average number of neighbors is

TABLE I. SIMULATION PARAMETERS.

Transmission Range	250 m	Topology Size	$1500 \times 300 \text{ m}^2$
Bandwidth of Radio Interface	2 Mbits/s	Traffic Type	CBR
Simulation time	900 sec	Packet Rate	4 packets/sec
Number of nodes	50	Packet size	512 Bytes
Number of source nodes	10, 20, 25, 30, 40	Mobile speed	0-20 m/s

around 7. The value of R is 30.

For the comparison 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, when a neighboring node first receives a RREQ packet, with probability T/M it rebroadcasts RREQ packet to its neighbors and with probability $1 - T/M$ it discards this RREQ packet. Table 2 provides a summary of the values of T , G , and p by varying the number of source nodes S .

TABLE II. SIMULATION VARIABLES.

variables # of sources, S	# of possible relay nodes, T	# 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

We compare three performance metrics for evaluations:

- *Packet delivery fraction*: The measured ratio between the number of data packets delivered to the destinations and the number of packets generated by all traffic sources. This evaluates the ability of the protocol to discover routes.
- *Average end-to-end delay of transferred data packets*: This includes all possible delays caused by buffering during route discovery, queuing at the interface-queue, retransmission delays at the medium access control layer, and propagation and transfer times.
- *Routing overhead (packets)*: The number of routing control packets per data packet delivered at the destination.

B. Simulation Results and Discussions

Figure 4 shows the performance of the network with different number of CBR sources. When the number of sources is less than 20, the network is not congested. All three routing protocols provide a high packet delivery fraction, small end-to-end delay and routing overhead. When the number of sources is above 20, the performance differences between LB-

Figure 4. Performance comparisons among AODV, LB-AODV, and GOSSIP1 routing protocols with different number of CBR sources (pause time = 500 sec).

AODV and the other two protocols become significant. LB-AODV outperforms the other two by providing a higher packet delivery fraction, a smaller end-to-end delay, and a significantly smaller amount of routing overhead.

Results in Figure 4(a) indicate that LB-AODV improves packet delivery fraction by 15% over the other schemes when the number of sources increases to 25. As traffic increases, the improvement is increased radically. This implies that when traffic load is high (i.e., more than 25 sources), most of the routes towards the gateway are congested by a lot of control and data packets. Therefore, contention and collision between

Figure 5. Performance comparisons among AODV, LB-AODV, and GOSSIP1 routing protocols with different pause time (number of CBR sources = 25).

neighbors increase exponentially, and thus AODV and GOSSIP1 routing schemes become less efficient.

Results in Figure 4(b) indicate that within a given end-to-end delay constraint, LB-AODV can support more traffic when compared with others. When the network is congested, the end-to-end delay of using LB-AODV is at least two times smaller than that of using other protocols.

Figure 4(c) shows that LB-AODV has much lower routing overhead when compared with AODV and GOSSIP1. If the number of source nodes is more than 25, the difference becomes significant. According to the Section II-C, the

maximum number of mobile nodes T that can relay the RREQ packet for each group is 30 (i.e., $M - S + (S/G)$) when the total number of groups G is 5, the number of sources S is 25, and the number of mobiles M is 50. On the other hand, when the number of groups G is equal to 1 (i.e., the original AODV), the maximum number of mobile nodes T that can relay the RREQ packet is 50. Since, in most on-demand routing schemes the loss of routing packets due to congestion may trigger the immediate generation of more routing packets, this explains the results for AODV and *GOSSIP1* that the network becomes highly congested with routing packets as the number of source nodes increases.

Figure 5 shows the overall performance with varying pause time (i.e., mobility). These results indicate that in a slightly congested network with number of sources equal to 25, LB-AODV maintains a higher packet delivery fraction (greater than 95%), a smaller end-to-end delay (less than 150 ms), and a lower normalized routing overhead when compared with the original AODV and gossip-based routing protocols.

Results in Figure 5(c) show LB-AODV incurs a lower routing control overhead than others. Recall that LB-AODV only allows a node to transmit a RREP packet in response to a RREQ packet when the node not only has a routing table to the gateway but can also maximize the balance index. Thus, the number of RREP packets transmitted to the source nodes is also much smaller. However, in the original AODV, all the intermediate nodes that have valid routing information to the gateway reply to all RREQ packets. This causes RREP packet flooding and wastes wireless resources due to unnecessary collisions in the link layer (i.e., IEEE 802.11). Our proposed LB-AODV is more efficient in the use of wireless resources than the original AODV in a mobile ad-hoc wireless access network environment. These results also imply that performance improvements correlate not only with guaranteeing network connectivity but also with the amount of routing load reduction.

IV. CONCLUSION

In this paper, we have proposed an extension of the ad-hoc on-demand routing protocol by incorporating the concept of load-balancing. Our proposed LB-AODV is well-suited for the mobile ad-hoc wireless access network environment where a gateway node provides access to the Internet. We have compared the performance of our proposed LB-AODV with both the original AODV and gossip-based routing protocols in different mobility and traffic scenarios. In scenarios with traffic congestion, LB-AODV significantly outperforms AODV and *GOSSIP1*. Simulation results show that LB-AODV delivers more data packets to the gateway and

decreases the end-to-end delay of packets delivered. In addition, LB-AODV can save valuable wireless resources by reducing the transmissions of routing control messages. Although we have presented our scheme based on the AODV routing protocol, the load-balancing concept can also be applied to other on-demand routing schemes.

To facilitate practical implementation of our proposed scheme, we are investigating techniques that provide good estimations of network size and topology in dynamic MANETs. We are also considering how our load-balancing concept can be incorporated in other on-demand routing protocols with different metrics (e.g., the least load or least power route).

ACKNOWLEDGMENT

This work was supported by a University of British Columbia Graduate Fellowship, and by the Canadian Natural Sciences and Engineering Research Council under grants RGPIN 261604-03 and RGPIN 44286-00.

REFERENCES

- [1] X. Hong, K. Xu, and M. Gerla, "Scalable Routing Protocols for Mobile Ad-hoc Networks," *IEEE Network*, vol. 16, pp. 11-21, July/Aug. 2002.
- [2] 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'01*, Helsinki, Finland, June 2001.
- [3] S.-T. Sheu and J. Chen, "A Novel Delay-Oriented Shortest Path Routing Protocol for Mobile Ad-hoc Networks," in *Proc. IEEE ICC'01*, Helsinki, Finland, June 2001.
- [4] J. Li, C. Blake, S. J. Douglas, D. Couto, H. Lee, and R. Morris, "Capacity of Ad-hoc Wireless Network," in *Proc. ACM MOBICOM'01*, Rome, Italy, Sept. 2001.
- [5] S. Roy and J. J. Garcia-Luna-Aceves, "Node-Centric Hybrid Routing for Ad-Hoc Wireless Extensions of the Internet," in *Proc. IEEE GLOBECOM'02*, Taipei, Taiwan, Nov. 2002.
- [6] C. E. Perkins, E. M. Royer, and S. R. Das, "Ad-hoc On-Demand Distance Vector (AODV) Routing," *IETF RFC 3561*, July 2003.
- [7] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-Based Ad-hoc Routing," in *Proc. IEEE INFOCOM'02*, New York City, New York, June 2002.
- [8] P. Hsiao, A. Hwang, H. Kung, and D. Vlah, "Load-Balancing Routing for Wireless Access Networks," in *Proc. IEEE INFOCOM'01*, Anchorage, Alaska, April 2001.
- [9] The Network Simulator - NS-2 notes and documentation and source code. Available from <http://www.isi.edu/nsnam/ns/>.
- [10] 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'98*, Dallas, Texas, Oct. 1998.
- [11] B. Tech "Development of WaveLAN, an ISM Band Wireless LAN," *AT&T Technical Journal*, pp. 27-37, July/Aug. 1993.
- [12] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, *ISO/IEC 8802-11; ANSI/IEEE Std 802.11*, Aug. 1999.