Network Architecture to Support QoS in Mobile Ad Hoc Networks

Lei Chen and Wendi Heinzelman
Department of Electrical and Computer Engineering
University of Rochester
Rochester, NY 14627
{*chenlei, wheinzel*}*@ece.rochester.edu*

Abstract

In this paper, we present a quality of service (QoS) architecture for supporting real-time data transmission in mobile ad hoc networks (MANETs). The QoS architecture includes a QoS transport layer, QoS routing, queue management and a priority MAC protocol. Through simulations, we find that the QoS architecture reduces packet delay and greatly improves the quality of real-time video streams in MANETs.

1. Introduction

Providing support for real-time data transmission is an important yet challenging goal for MANETs. Much research has been done in each network layer to support real-time data transmission. Various routing protocols have been proposed (e.g., [1] [2]) that either provide admission control or find a path with large enough bandwidth to support a given request. Packet queue management in the link layer has been presented in [3] and [4], where [3] compares two different queue management algorithms and [4] evaluates the performance of different queue scheduling algorithms with the DSR and GPSR routing protocols. Priority scheduling in the MAC layer has been studied in [5]. While these protocols improve the support for real-time data transmission in MANETs, data communication is the result of each network layer's effort; thus, the cooperation of all network layers is needed to provide QoS support. A service differentiation model has been presented in [6], and QoS models have been discussed in [7]. However, these designs are not comprehensive enough to include all the networking layers. Therefore, we propose a QoS architecture that extends from the application layer to the MAC layer to support real-time data transmission. This QoS architecture is described in the following section.

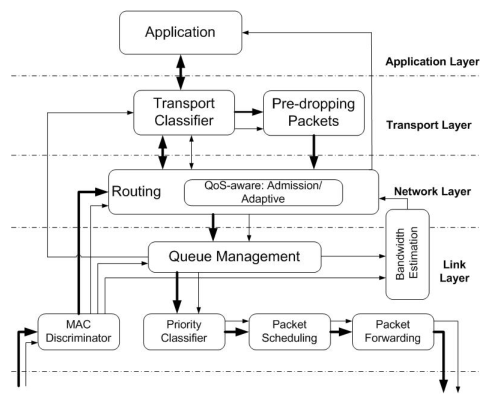

Figure 1. QoS architecture.

2. QoS architecture

In Figure 1, we show our proposed QoS architecture, which includes all networking layers from the application layer to the MAC layer. The bold lines indicate the flow of data packets and the narrow lines indicate the flow of control packets. Each layer's features are detailed below.

2.1. Application layer

Applications can be categorized into real-time and non-real-time applications based on their sensitivity to packet delay. Real-time applications have strict requirements on the packet delay. Therefore, packet retransmission is not allowed. The applications that fit into this category are on-line live movies and video conferencing. Many video compression technologies, such as MPEG-4, H.263, and multiple-description coding, can compress video with different coding rates to meet different channel conditions. In addition, most of these compression schemes have error resilience features to recover the video frame, if some packets are lost. Thus, choosing the right coding rate to compress the video is important, and some reasonable packet loss is acceptable. On the other hand, for non-real-time applications such as Email and FTP, packet delay is not a big issue, and packet delivery is guaranteed by explicit acknowledgements in the transport layer. The network should be designed to meet the different packet delay requirements of these two types of applications.

2.2. Transport layer

UDP and TCP are two transport layer protocols widely used in wired networks. UDP has no congestion control scheme to react to network congestion. Applications that use UDP as the underlying transport protocol to transmit packets can easily overwhelm the network with data, which results in a considerable amount of wasted energy and bandwidth in transmitting packets that will be dropped due to congestion. Therefore, some pre-dropping of UDP packets should be investigated to react to congestion [8]. TCP has an inherent congestion control scheme, so congestion control is not a problem. However, TCP's performance should be optimized to adjust the TCP window, which requires feedback information from the lower network layers [9]. Therefore, some information from the packet queue and the routing layer should be sent to the transport layer for performance optimization.

2.3. Network layer

To support QoS, the routing protocol should have an embedded scheme such as call admission or adaptive feedback that is designed to support QoS. At the same time, non-QoS-aware routing that is targeted at finding a feasible path should be offered as well. For QoS-aware routing, information about the current network status is provided to the application for performance optimization. Also, the routing layer should get enough channel information from the lower layers so that the admission/adaptive scheme can be performed based on the network status. Therefore, two cross-layer designs should be implemented in QoS-aware routing. One is to obtain the network resource information from lower layers, and the other is to send the network status to the applications. To offer QoS to the applications, resource reservation should be incorporated. An RSVP-type signaling scheme [10] is not desirable in MANETs due to its high overhead. Therefore, in-band and soft resource reservation (i.e., best effort rather than guaranteed reservations) should be done during the route discovery phase and during route maintenance. The transmissions that occur between the break down of old routes and the set up of new routes will severely affect the QoS provided by the network. Therefore, some prediction of route breaks should be incorporated. Overall, QoS-aware routing should have the following features that traditional routing does not support:

- obtain resource information from lower layers;
- offer bandwidth information to applications;
- incorporate resource reservation schemes; and
- predict route breaks.

2.4. Link layer

The link layer needs to discriminate the different priority packets and schedule packet delivery according to priority levels. The service differentiation should be completed in the packet queue through queue management and in the MAC layer through a MAC discriminator and priority classifier.

Queue Management: The aim of queue management is to schedule the different priority packets. Real-time data should have higher priority to be sent to the channel compared with packets such as FTP and Email. Therefore, real-time data will be put in front of the non-real-time data in the packet queue. When the network is congested, the last packet in the packet queue will be dropped. Therefore, incorporating queue management will reduce the possibility that real-time packets are dropped in the packet queue when the network is congested. Thus, the delay of real-time application data packets can be reduced and the packet delivery ratio can be improved. Also, the packets whose delay has already exceeded the applications' requirement should be eliminated from the packet queue before transmission to save the transmission of packets that will be useless to the receiver. If different flows go through the same host, it is easier to do the priority regulation in the packet queue than in the MAC layer.

MAC Discriminator: The main function of the MAC discriminator is to differentiate data packets and control packets that arrive from the wireless channel. Data packets are sent to the network layer; ARP (address resolution protocol) packets go to the queue directly; MAC packets, such as the RTS, CTS, and ACK packets used in IEEE 802.11, stay in the MAC layer; and the bandwidth estimation control packets are sent to the bandwidth estimation module for use in the routing layer's admission/adaptive scheme.

Priority Classifier and Packet Scheduler: To offer service differentiation in a distributed ad hoc network, real-time packets should be granted higher priority to capture the channel. The priority classifier differentiates the different data packets that arrive from the packet queue and directs the packet scheduler to schedule the packet delivery based on the priority level of the current packet.

2.5. Bandwidth estimation

In a distributed ad hoc network, a host's available bandwidth is not only decided by the raw channel bandwidth, but also by its neighbor's bandwidth usage and interference caused by other sources, each of which reduces a host's available bandwidth for transmitting data. Therefore, applications cannot properly optimize their coding rate without knowledge of the status of the entire network. Thus, bandwidth estimation is a fundamental function that is needed to provide QoS in MANETs. However, bandwidth estimation is extremely difficult, because each host has imprecise knowledge of the network status and links change dynamically. Therefore, an effective bandwidth estimation scheme is highly desirable. Bandwidth estimation can be done using various methods; for example, in [1] bandwidth estimation is a cross-layer design of the routing and MAC layers, and in [2], the available bandwidth is estimated in the MAC layer and is sent to the routing layer for admission control. Therefore, bandwidth estimation can be performed in several different network layers, as shown in Figure 1.

3. Simulations

We built a simplified network model to support real-time data transmission. In the simplified network model, the following designs are incorporated.

- Priority packet scheduling scheme: The packets with high priority (e.g., real-time data packets, routing packets and ARP packets) are always put in front of the non-real-time data packets. The packets in the tail of the queue are dropped when congestion occurs.

- Bandwidth estimation-based QoS-aware routing: Each host periodically estimates its own bandwidth use with MAC layer bandwidth estimation, and this information is disseminated to the host's two-hop neighbors through "Hello" packets. Each host's available bandwidth is estimated based on the bandwidth used by itself as well as each of its one-hop and two-hop neighbors. Either an admission control scheme is used during route discovery, or adaptive feedback is embedded in the route reply packets. This procedure, which is an extension to AODV, is detailed in [11].

- Enhanced IEEE 802.11: AIFS (Arbitration Interframe Space) and different window sizes [5] are applied to help high priority packets capture the channel. The MAC layer assigns shorter AIFS and smaller window sizes to real-time data packets.

Table 1. Architectures used in simulations.

Architecture	Routing	Queue	MAC
No QoS	AODV	FIFO	IEEE 802.11
QoS1	QoS-aware	Priority Queue	IEEE 802.11
QoS2	QoS-aware	Priority Queue	Enhanced IEEE 802.11

To test the performance of our simple QoS architecture, we simulate an ad hoc network using the ns-2 simulator. The H.263 TMN simulator is used to show the video quality improvement using a QoS architecture in ad hoc networks. In our simulations, 50 static nodes are randomly placed within a 1000 m by 1000 m area, and we simulate twenty different random scenarios. The packet size used in our simulations is 500 bytes and the raw channel bandwidth is 2 Mbps. Three source-destination pairs are randomly chosen to simultaneously transmit non-real-time data packets, and one randomly chosen real-time data flow joins in after 10 seconds. The total simulation time is 200 seconds. UDP is used as the underlying transport layer protocol for both the real-time and the non-real-time streams.

The three different network architectures shown in Table 1 are tested. No QoS uses the regular AODV routing, a FIFO queue, and the IEEE 802.11 MAC. QoS1 uses QoS-aware routing with adaptive feedback, priority packet scheduling, and the IEEE 802.11 MAC. QoS2 uses QoS-aware routing with adaptive feedback, priority packet scheduling, and the IEEE 802.11 MAC with AIFS and smaller window sizes for high priority data packets.

All flows initially attempt to send at the same rate. The real-time application can adjust its sending rate according to the network status if QoS-aware routing is used and the available bandwidth information is sent to the application by the cross-layer design. Figure 2 shows the real-time data flow's delay. We can see that using QoS-aware routing and priority queue management reduces the average packet delay, and incorporating differentiated MAC further decreases the delay.

We randomly choose one of the twenty scenarios to test one real-time data flow and three non-real-time

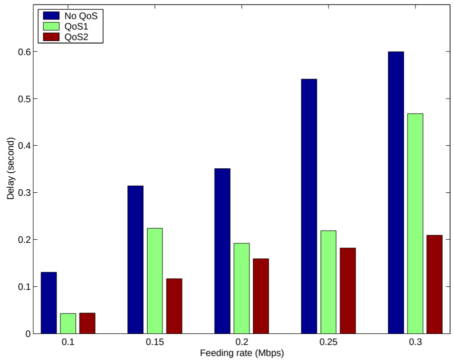

Figure 2. Real-time data flow's average packet delay.

Figure 3. Video frame using the QoS2 architecture.

data flows. The three non-real-time data flows transmit data at 80 kpbs, and the real-time application begins to send a video stream at 300 kbps ten seconds later. Using the QoS2 architecture results in the video frame quality shown in Figure 3; using the No QoS architecture results in the video frame quality shown in Figure 4. Using adaptive feedback, the application compresses the video according to the network status, which avoids overwhelming the network, so congestion can be avoided. Therefore, the packet delivery rate is increased, which helps to improve the video quality.

4. Conclusion

We present a network architecture that supports QoS in MANETs, and we simulate a simplified version of this QoS architecture. Our simulation results show that video quality can be greatly improved and average packet delay can be significantly decreased using our QoS architecture. However, the performance of multiple different priority streams has not been presented. We will address this issue in our future work.

Figure 4. Video frame using the No QoS architecture.

5. References

[1] Q. Xue and A. Ganz, "Ad hoc QoS On-demand Routing (AQOR) in Mobile Ad hoc Networks," in *Journal of Parallel and Distributed Computing (JPDC)*, February 2003.

[2] Y. Ge, T. Kunz, and L. Lamont, "Quality of Service Routingin Ad-hoc Networks Using OLSR," in *Proceeding of the 36th Hawii International Conference on System Science*, 2003.

[3] H. Xiao, K.C. Chua, and W.S.A. Lo, "On Service Prioritization in Mobile Ad-hoc Networks," in *Proceeding of ICC*, 2001.

[4] B-G. Chun and M. Baker, "Evaluation of Packet Scheduling Algorithm in Mobile Ad Hoc Networks," in *Mobile Computing and Communications Review*, 2002.

[5] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e Wireless LAN for Quality of Service," in *European Wireless*, 2002.

[6] G. Ahn, A. Campbell, A. Veres, and L.Sun, "SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks," in *Proceedings of IEEE INFOCOM*, 2002.

[7] H. Xiao, K.C. Chua, and W.K.G. Seah, "A Quality of Service Model for Ad Hoc Wireless Networks," in *Handbook of Ad Hoc Wireless Networks*, FL USA, 2002, CRC.

[8] L. Chen and W. Heinzelman, "End-to-End Congestion Control for Best-effort Transmission," in *Proceedings of Wireless Networking Symposium*, October 2003.

[9] K. Chandran, S. Raghunathan, S. Venkatesan, and R. Prakash, "A Feedback-based Scheme for Improving TCP Performance in Ad Hoc Wireless Networks," in *IEEE Personal Communications Magazine*, Febuarary 2001.

[10] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A New Resource ReSerVation Protocol," in *IEEE Network*, September 1993.

[11] L.Chen and W. Heinzelman, "QoS-aware Routing Based on Bandwidth Estimation in Mobile Ad Hoc Networks," in *Technical Report of University of Rochester*, 2003.