A QoS Routing and Admission Control Scheme for 802.11 Ad Hoc Networks

Lin Luo, Marco Gruteser, Hang Liu and Dipankar Raychaudhuri
WINLAB, Rutgers University
671 Route 1 South, New Brunswick, NJ 08902
{clarylin, gruteser, hliu, ray}@winlab.rutgers.edu

King Huang, and Shiwen Chen
PSDU, Panasonic Technologies Company
2 Research Way, Princeton, NJ 07746
{king, schen}@psdu.panasonic.com

ABSTRACT

This paper presents an admission control mechanism for multi-rate wireless ad hoc networks. Admission control depends on precise estimates of bandwidth available in the network and the bandwidth required by a new flow. Estimating these parameters in wireless ad hoc networks is challenging due to the shared and open nature of the wireless channel. Available bandwidth can only be determined by also considering interference at neighboring nodes. Also, due to self-interference of flows the required bandwidth of a flow varies for each link of a route. The proposed admission control mechanisms is integrated with a hop-by-hop ad hoc routing protocol, thus enabling it to identify alternate routes if the shortest path is congested. Each node measures available channel bandwidth through passive monitoring of the channel. The mechanism improves estimation accuracy by using a formula that considers possible spatial reuse from parallel transmissions. The protocol also uses temporal accounting to enable bandwidth estimation across links using different bit-rates. Simulation results support that the admission control mechanism can effectively control the traffic load and that considering parallel transmission leads to improved bandwidth estimation accuracy. The admission control mechanism can admit more traffic while maintaining QoS.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: design studies; C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*wireless communication*

General Terms

design, performance

Keywords

QoS routing, admission control, parallel transmission

1. INTRODUCTION

The higher data rates supported by short-range wireless networking standards enable a variety of new media streaming and distribution applications. IEEE 802.11-based wireless networks allow streaming of audio and video content between home entertainment devices. Wireless networks can also reduce the cost of deploying arrays of surveillance cameras. The range and reliability of such systems can be extended through multi-hop communication if the direct link to a base station fails. Since real-time media streams require low delays and packet loss rates, these applications benefit from Quality of Service (QoS) mechanisms such as admission control. Admission control prevents the network from reaching congestion by rejecting new media streams if insufficient bandwidth is available.

Providing admission control in wireless networks is particularly challenging because of the shared and open nature of single channel wireless communications. Unlike wired networks, the available bandwidth on a communication link may change due to mobility or outside interference. While these issues prevent wireless QoS solutions from providing strong bandwidth guarantees, wireless quality-of-service mechanisms can still increase the reliability of the communication link. In wireless networks, each node on the channel may also have a different view of channel utilization based on their unique position in space. In addition, wireless transmission also contend with transmissions outside the direct communication range. The contention or carrier-sense range is typically much larger than the transmission range, making it difficult to determine whether a newly admitted flow affects existing flows at neighboring nodes.

This paper presents an admission control mechanism integrated with ad hoc routing. Wireless networks often provide multiple possible routes between a source destination pair. Thus, integrating admission control into the routing protocol enables the identification of alternate routes if the shortest path is congested. Most closely related to this work is the Contention-Aware Admission Control Protocol independently developed by Yang and Kravets [10]. This paper builds on CACP with the following key contributions:

- A mechanism for accurate admission control decisions in a multi-rate environment (as in a typical 802.11 network). In multi-rate networks, the mechanisms must realize that the link utilization required by a flow is a link-oriented concept, depending on the bitrates and position of the link.
- Protocol extensions to exchange topology information needed for admission in hop-by-hop routing protocols such as AODV or LUNAR. Earlier work, such as CACP used source routing

protocols, where the route information is contained in the packet headers.

- A formula considering channel reuse due to parallel transmissions for more accurate channel utilization estimation. It is based on passive monitoring with dual carrier-sensing thresholds as proposed in CACP.

The rest of the paper is organized as follows: in section 2, we explain the challenges to perform admission control in wireless networks. Section 3 presents our solutions of predicting the link utilization of the requesting flow, estimating channel availability and making admission decision. Then in section 4, the proposed scheme and its implementation are described in detail. Section 5 shows the simulation results to demonstrate the effectiveness of our proposed scheme. Finally, section 6 concludes our work.

2. ADMISSION CONTROL IN WIRELESS AD HOC NETWORKS

Admission control allows a new data flow onto the network if the available network bandwidth is greater than the bandwidth required by the flow. Thus, making an admission decision requires an estimate of the available network bandwidth and an estimate of the bandwidth required by the flow. In a network, admission control must ensure sufficient available bandwidth on the bottleneck link of the route, because congestion may differ on each link. Since single-channel wireless links are not isolated from each other, admission control also has to ensure that the admission of a flow does not interfere with existing flows on nearby links. Assuming symmetric channels and identical radio configurations, a node's transmission could interfere with flows passing through any node in the carrier-sensing range of the respective node. We refer to these nodes as *carrier sensing nodes (CSN)*.

The admission control mechanisms must ensure that each of the CSNs has sufficient available bandwidth to accommodate the new flow. Available bandwidth, however, is link-dependent due to differences in bitrate and packet error rate (the same node can have different bandwidth available for different communication links). Therefore, it is more meaningful to measure a node's channel availability in terms of time, which is independent of these factors. Specifically, we define a node's *channel availability* as the fraction of the channel available for transmissions without interfering with existing flows on its CSNs.

The required link utilization of a flow depends on the data rate and packet size used by the flow in addition to the bitrate on the link. However, due to self-interference of flows in wireless networks, the link utilization is also affected by the fraction of channel the flow consumes on other nearby links.

2.1 Related Work

Much work has been done in providing QoS support in wired and wireless networks. IntServ/RSVP and DiffServ[2] are two solutions for QoS provisioning on the Internet, but they are not suitable for wireless ad hoc networks [3]. They can not address the challenges of estimating the available resources and calculating the resources required by the flow due to shared nature of wireless channels.

Some QoS solutions have been proposed for wireless ad hoc networks. In [4], admission control is mentioned as a necessary component to support QoS in wireless ad hoc networks, but no specific algorithms were designed. In SWAN[5], the admission controller promiscuously listens to all packet transmissions within its transmission range to gather information of bandwidth and congestion. Admission decision is based on the bandwidth measured along the path of communication by sending a probe message. Probing introduces a lot of overhead and may not be able to determine an accurate value if packet loss occurs. Besides, SWAN does not consider the fact that nodes could interfere with each other even though they may not communicate directly. The same problem exists in [6] and [7]. [8] presents a resource reservation-based routing and signalling algorithm, ad hoc QoS on-demand routing (AQOR). It aims to provide end-to-end QoS support in terms of bandwidth and end-to-end delay. In AQOR, every node is required to periodically send out a "Hello" packet, announcing its existence and traffic information to its neighbors. Based on the exchanged information, every node calculates the available bandwidth and bandwidth consumption, taking into account self traffic, neighborhood traffic and boundary traffic. However, AQOR does not give enough attention to the fact that, when making admission decision, a node must consider not only local resources but also the resources at all the contending nodes within its CSR. [10] is most related to our work. In this work, the authors proposed the Contention-aware Admission Control Protocol (CACP) to support QoS in ad hoc networks. CACP is the first work to introduce the concept of c-neighborhood available bandwidth, which refers to the available bandwidth at a node's CSNs. It requires that a node must have enough local and c-neighborhood available bandwidth to successfully admit a flow. CACP depends on source routing to build the "c-neighbor set" to calculate bandwidth consumption. We build on CACP and develop an admission control mechanisms that can operate in a multi-rate environment and provides more accurate available bandwidth estimation by considering parallel transmissions. We also develop a protocol for sharing topology in common hop-by-hop routing protocols instead of source routing.

3. ADMISSION CONTROL MECHANISM

The admission control mechanism comprises link utilization prediction, channel availability estimation, and routing protocol extensions to share necessary information between nodes.

3.1 Prediction of Link Utilization of a Flow

The required link utilization of a flow can be predicted by computing the medium access control overhead and by estimating the self-interference component based on the node's position in the route. To enable this prediction the average packet rate and packet size must be known. These can be either provided by the application generating the flow or estimated from the first packets generated for this flow. In the following derivation, we also assume a perfect channel and do not consider transmission failures due to channel errors, but the formula can be extended with a packet error rate component on each link if the average packet error rate can be measured by the nodes.

3.1.1 Link Utilization Requirement of the Source

For IEEE 802.11 MAC using RTS-CTS-DATA-ACK handshake, as shown in Fig. 1, per-hop occupation time of a data packet, T_{occup}, can be expressed as:

$$
\begin{aligned}
T_{occup} = & T_{rts} + T_{cts} + \frac{L}{B} + T_{PLCP} + T_{ack} + \\
& 3T_{sifs} + T_{difs} + T_{backoff}
\end{aligned} \quad (1)
$$

In (1), L is the size of the data packet including MAC header; T_{PLCP}, T_{rts}, T_{cts} and T_{ack} represent the time for transmitting *PLCP* header, *RTS*, *CTS* and *ACK* packets respectively. T_{sifs} and T_{difs} denote the inter-frame spaces SIFS and DIFS, respectively, which are defined in the IEEE 802.11 standard. B is the link rate

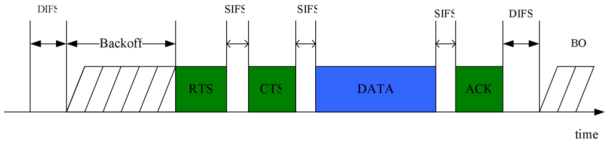


Figure 1: Operation of Four-Way Handshake in the 802.11 DCF

used by the source node. $T_{backoff}$ denotes the backoff time before the transmission, which can be expressed as $\frac{CWMin}{2} \cdot SlotTime$. Since our scheme is designed to control the traffic load and provide QoS guarantees to admitted flows, transmission failures due to collisions are expected to be negligible. Thus, the contention window (CW) can be assumed to mostly remain at $CWMin$. Since backoff slots are uniformly distributed in $[0, CW]$, the expected slot number is $\frac{CWMin}{2}$ and the backoff time can be obtained by multiplying with the slot time.

In the expression of T_{occup}, only the $\frac{L}{B}$ term is link-bitrate dependent. By defining $T_{data} = \frac{L}{B}$,

$$T_{occup} = T_{data} + T_{oh} \tag{2}$$

where T_{oh} is the sum of rate independent terms in (1).

If the application at the source node generates *R* packets per second, then its requirement or link utilization for one-hop transmission ρ_{req} can be expressed as:

$$\rho_{req} = R \times (T_{data} + T_{oh}) \tag{3}$$

Since T_{data} is dependent on the link rate B, ρ_{req} also depends on B. However, when the source node specifies ρ_{req}, it only knows its own link rate. On a link with a different rate, the link utilization would differ. Therefore, we need to provide necessary information to other nodes, letting them be able to calculate the link utilization locally. We will describe it in the protocol implementation section.

3.1.2 Estimating flow self-interference

The total link utilization required by the requesting flow on a link depends on its position on the path, since nodes on the same path will contend for the channel with each other, which we call self-interference or intra-flow contention. Consider the sample path in Fig. 2. The dashed circle represents A's carrier-sensing range and the solid arrows stand for the requesting flow from C1 to C4. For the admitting node A, nodes N1, N2, N3 and N4 are all its CSNs. Transmissions from each of these nodes as well as from node A itself add to the link utilization on the link $A - N3$. Assume that the link rates on $N1-N2$, $N2-A$, $A-N3$, $N3-N4$ and $N4-C3$ are B_1, B_2, B_3, B_4 and B_5, respectively. The aggregate link utilization required by the requesting flow can then be calculated as $\sum_{i=1}^{5} R \cdot (\frac{L}{B_i} + T_{oh})$.

In general, aggregate link utilization required by the requesting flow on a node's outbound link can be formulated as follows:

$$\rho_{aggr} = \sum_{i=1}^{N_{cont}+1} R \cdot (\frac{L}{B_i} + T_{oh}) \tag{4}$$

where L is the packet size and R is the packet sending rate. N_{cont} denotes the number of the node's CSNs on the path excluding the destination, since the destination only passively receives the packets.

N_{cont} is hard to obtain in mobile ad hoc networks where the topology is not known in advance. In a single-rate network, we can approximate N_{cont} of each node with the number of non-destination

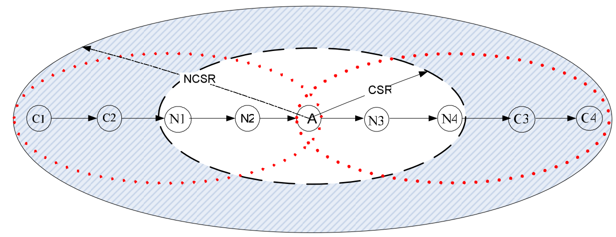


Figure 2: Example of Calculating Bandwidth Consumption

nodes within its $[\frac{CSR}{TxR}]$ hops. For example, ns-2 defines the carrier-sensing range in IEEE 802.11 as approximately twice the transmission range at 2Mbps bit rate. Thus, we can use the number of 2-hop neighbors on the path to approximate N_{cont} in a 2Mbps network. In Fig.2, if assuming all the links work at 2Mbps bit rate, then N_{cont} of node A is 4, because N1, N2, N3 and N4 are within its two hops and none of them is the destination. However, N_{cont} of N4 is only three. Although there are 4 nodes within its two hops, one of them (C4) is the destination. This approximation, however, can not necessarily apply to a multi-rate environment. For example, the popular 802.11b physical (PHY) layer provides 1, 2, 5.5 and 11Mbps. If each link in the network works at either 1Mbps or 2Mbps, we can still use the 2-hop approximation, since the transmission ranges at the two bitrates are both about one half of the carrier-sensing range. To estimate N_{cont} in an environment with more rates, some link-bitrate information could be distributed in the route discovery. For example, an intuitive way is to let each node appends its bitrate in the route packets before forwarding them. Given the correspondence between bitrate and transmission range, N_{cont} can be easily estimated. This approach can be optimized if we know some general bitrate information, such as which rates are being used in the network.

3.2 Estimation of Channel Availability

We estimate channel availability through passive monitoring at each node with a lowered carrier-sensing threshold, as suggested in [10]. To improve estimation accuracy, our approach measures busy time using two thresholds, to be able to estimate and subtract the amount of possible parallel transmissions.

First, local channel busy time (T_{busy}^{local}) can be measured by passively monitoring the transmission activity in the regular carrier-sensing range. Here we define the channel busy time as the total time that a node is transmitting, receiving or has sensed carrier signals. When the signal strength is higher than the *carrier-sensing threshold*, the channel is assumed busy. To get the information of channel busy time at its CSNs (T_{busy}^{csn}), a node extends its measurement range to enclose the carrier-sensing ranges of all its CSNs. Figure 2 illustrates this relationship. This extended range is called *neighbor-carrier-sensing range* (NCSR) and can be implemented using a second lower *neighbor-carrier-sensing threshold*. If the signal strength is higher than this threshold the channel is assumed busy at the sensing node's CSNs. Note, however, that this estimate of channel busy time at a node's CSNs is overly conservative due to the assumption that any transmission activity in a node's neighbor-carrier-sensing range consumes the bandwidth at all its CSNs and that no transmision can occur in parallel.

During the measurement interval T_p, let T_{busy}^{local} be the local busy time sensed by the admitting node when the signal strength is above *carrier-sensing threshold* and T_{busy}^{csn} be the busy time when the signal strength is larger than *neighbor-carrier-sensing threshold*. With these two measurements of channel busy time, $\frac{T_{busy}}{T_p}$ is used to es-

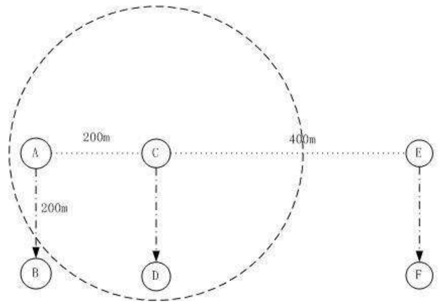


Figure 3: Illustration of Parallel Transmission

timate the local channel utilization ρ_{local} and channel utilization at the node's CSNs ρ_{csn} as follows:

$$\rho_{local} = \frac{T_{busy}^{local}}{T_p} \text{ and } \rho_{csn} = \frac{T_{busy}^{csn}}{T_p}$$

Since T_{busy}^{csn} is measured with a lower threshold, it fully contains T_{busy}^{local}. The idle fraction of the channel can thus be calculated as $\rho_{idle} = 1 - \rho_{csn}$. This already accounts for interference at the CSNs, meaning that only flows could be admitted that would not interfere with other flows at the CSNs.

Besides the idle channel fraction, parallel transmission can generate some extra channel share available to the flow. Consider Fig 3. If flow EF were admitted, the total channel busy time measured by C would be $T_{AB} + T_{CD} + T_{EF} - T_{overlap}$, where $T_{overlap}$ is the fraction of A and E transmitting at the same time. Dividing by the measurement interval T_p, we can calculate the channel utilization as $\rho_{AB} + \rho_{CD} + \rho_{EF} - \rho_{overlap}$. Since each individual utilization component can be interpreted as the probability that a node transmits at a given time, we can approximate the amount of overlap by assuming that the transmission of AB and EF are independent of each other. This leads to $\rho_{overlap} = \rho_{AB} \cdot \rho_{EF}$. Thus, channel availability for flow EF is $1 - (\rho_{AB} + \rho_{CD}) + \rho_{overlap}$, which represents the fraction that node E can transmit without interfering with AB and CD.

3.3 Admission Decision with Parallel Transmission Consideration

When admitting a new flow, we estimate how much the new flow can transmit in parallel with existing nodes outside the carrier sensing range to obtain a more accurate estimate of the required utilization. Parallel transmission can be estimated based on the two time measurements and link requirement of the flow.

Since T_{busy}^{csn} is measured with a lower threshold, it fully contains T_{busy}^{local}. $T_{busy}^{csn} - T_{busy}^{local}$ is thus an estimate of the amount of channel busy time contributed by the transmissions outside the admitting node's carrier-sensing range. In Fig. 2, for example, if A is the admitting node, $T_{busy}^{csn} - T_{busy}^{local}$ then reflects the activities within the light shaded area. Since transmissions outside a node's carrier-sensing range (CSR) do not interfere with transmissions from the node itself, the parallel transmission part is

$$\rho_{overlap} = \frac{T_{busy}^{csn} - T_{busy}^{local}}{T_p} \times R \cdot (T_{oh} + \frac{L}{B_a})$$

where $R \cdot (T_{oh} + \frac{L}{B_a})$ represent A's transmission and B_a is the link rate of node A.

If the flow were admitted, link utilization at the admitting node A updates to

$$\rho_{local}^{u} = \rho_{local} + \rho_{aggr}$$

Meanwhile, the total link utilization at node A's carrier-sensing neighbors (CSNs) updates to

$$\rho_{csn}^{u} = \rho_{csn} + \rho_{aggr} - \rho_{overlap}$$

To avoid congestion, the admitting node must ensure $\rho_{local}^{u} \leq 1$ and $\rho_{csn}^{u} \leq 1$. For more conservative admission control the utilization limit can also be chosen smaller than one—we use a value of one here for simplicity. We note that the second condition is more stringent than the first one. Therefore, when a node makes admission decision, it only needs to check if $\rho_{csn}^{u} \leq 1$ can be satisfied, that is,

$$\rho_{csn} + \rho_{aggr} - \rho_{overlap} \leq 1$$

which implies

$$\rho_{aggr} \leq 1 - \rho_{csn} + \rho_{overlap}, \text{ i.e,}$$

$$\rho_{aggr} \leq 1 - \frac{T_{busy}^{csn}}{T_p} + \frac{T_{busy}^{csn} - T_{busy}^{local}}{T_p} \times R \cdot (T_{oh} + \frac{L}{B_a}) \quad (5)$$

where the left side $\rho_{aggr} = \sum_{i=1}^{N_{cont}+1} R \cdot (\frac{L}{B_i} + T_{oh})$ is the aggregate link utilization required by the flow, while the right side represents the fraction of the channel available to the flow.

In a network where all the links use the same rate, equation (5) can be simplified to:

$$(N_{cont} + 1)\rho_{req} \leq 1 - \frac{T_{busy}^{csn}}{T_p} + \frac{T_{busy}^{csn} - T_{busy}^{local}}{T_p} \times \rho_{req} \quad (6)$$

where ρ_{req} is calculated in (3)

This inequality must be verified at each node during the route establishment phase. The following section describes the details of this proecess.

4. PROTOCOL IMPLEMENTATION

The proposed QoS scheme consists of four parts: route discovery, distributed admission control, channel reservation and QoS violation recovery. We discuss the protocol implementation in the context of the LUNAR routing protocol [11]. We have chosen LUNAR for the implementation of the mechanisms due to its simplicity and small code base. The concepts under discussion, however, can also be applied to other hop-by-hop routing protocols such as AODV. For ease of explanation, we only describe the implementation for a network with only 1Mbps and 2Mbps bitrates. However, it can be easily extended for the network with more rates by distributing more link-rate information to the nodes. To keep track of flow information, a *Flow Table* is maintained at every node in the network. Each entry of this table describes a flow passing through the node, including the source, destination, reserved channel and status. All entries are maintained as soft state, so that they are automatically deleted if the topology changes. The status can be one of the three values: requesting, reserved and activated. If multiple flows between the same source-destination pair need to be supported, an additional identifier can be added into the *Flow Table* entry to uniquely represent a data flow.

4.1 Route Discovery and Admission Control

We add several fields to the routing protocol to distribute topology and flow requirement information to each node that must make an admission decision. Upon receiving an ARP request, the source node broadcasts a resolution request (RREQ) to its neighbors. Besides the original fields defined in LUNAR, we add four more fields to the RREQ message to facilitate QoS provisioning. "Flow ID", associated with source and destination, uniquely identifies a flow.

"$SendingRate$" states packet sending rate R of the flow. We also include the packet size in the RREQ, denoted as "L". L and R characterize the traffic of the flow. "$B_{upstream}$" field specifies the link rate used by the previous hop, which is 0 at the source node since it is the first hop. $B_{upstream}$ is used by the next-hop node to calculate the aggregate link utilization required by the requesting flow. When receiving the RREQ, each intermediate node calculates the aggregate link utilization introduced by the transmissions over last two hops and the next hop. Link utilization of the immediate previous hop is calculated as $R \cdot (T_{oh} + \frac{L}{B_{up}})$, where B_{up} denotes the link rate. Link utilization of the second previous hop is $R \cdot (T_{oh} + \frac{L}{B_{upstream}})$, where $B_{upstream}$ is obtained from the received RREQ. Plus its own transmission on the next hop, the node can calculate the aggregate channel utilization as $\rho^{up}_{aggr} = R \cdot (T_{oh} + \frac{L}{B_{up}}) + R \cdot (T_{oh} + \frac{L}{B_{upstream}}) + R \cdot (T_{oh} + \frac{L}{B_{down}})$, where B_{down} is the link rate which the node will use on the next hop. This aggregate link utilization of the flow is put into (5) to check the channel availability and to make the admission decision. If the requirement can be met, the intermediate node re-broadcasts the RREQ message. Also it inserts an entry in the *Flow Table*, recording source, destination and ID of the flow, filling in the "*reserved channel*" field with ρ^{up}_{aggr} and setting the "status" to "*Requesting*". If the requirement can not be satisfied, the node discards the RREQ and does nothing further. Admission control performed in this phase is partial, because the admitting node can only obtain the information of its upstream contending nodes. Although it underestimates the aggregate channel utilization introduced by the flow, it is useful as the first pass to cheaply weed out certain non-qualified routes.

Upon receiving the resolution request, the intended destination sends back a resolution reply (RREP) if the channel availability is large enough. Note that multiple copies of RREQ might arrive along different paths. To increase the possibility of discovering a qualifying path, the destination sends back a RREP for each copy of the RREQ. We add the same four fields in the RREP as in RREQ. At each forwarding node, full admission control is performed. We call it "full" because in this phase the total link utilization required by the flow can be calculated. Each forwarding node calculates the aggregate utilization over the previous two hops the same way as in RREQ phase. Note that the "previous" side in the RREP phase is the "succeeding" side in the "RREQ" phase. Therefore, summing up the aggregate link utilizations calculated during the RREQ and RREP phase, an admitting node can get total channel utilization ρ_{aggr} of the requesting flow. ρ_{aggr} is inserted to (5) to make the admission decision. If it succeeds at a node, the RREP is forwarded to the next hop. At the same time, a soft reservation is setup at the node. The "*reserved channel*" field of the corresponding *Flow Table* entry is updated with ρ_{aggr} and the status becomes "*Reserved*".

When the first data packet of a flow arrives, the node update the status of the corresponding entry to "*activated*". The "*activated*" status means the bandwidth is being actually used.

When a channel reservation exists at a given node, the reserved utilization ($\rho_{resv} = \rho_{aggr}$) is no longer available to other flows and thus the idle fraction of the channel becomes $1 - \rho_{local} - \rho_{resv}$.

Although a node makes a reservation during the RREP phase, the route may not be successful due to any of a variety of reasons such as link failure or the source deciding to use a different route [8]. Therefore, reservations are maintained as soft state and are deleted when the associated timer expires. Specifically, each entry is associated with three timeout values: *route-reply timeout*, *data-start timeout* and *activity timeout*.The *route reply timeout* is used to remove a entry from the *Flow Table* if a node only receives a route request, but does not receive a route reply in time. The *data start timeout* expires if a node has received a route reply for a reservation, but has not yet received data from the source within this period. The *data timeout* is used to keep track of flows which have transmitted data, but this timeout value has elapsed since the last data packet was forwarded. This can occur, for example, due to node movement so that the currently used route fails and the source needs to use another route.

4.2 QoS Violation Detection and Recovery

We exploit the periodic route refresh mechanism used in LUNAR to provide an implicit QoS violation detection and recovery mechanism. If a node can no longer meet its QoS commitments made to existing flows, the admission control mechanism will reject or find alternate routes for some of these flows during the next route refresh (by default every 3 seconds). This means that when a new route request is received, a node should first delete any existing reservation with the same flow ID, as to prevent double counting of the bandwidth requested by this flow.

Refreshing path reservations creates a challenge in providing path stability. A flow should switches paths only when the current path can no longer meet the flow requirement. Unnecessary changes may lead to changes to other existing flows with possibly some flows alternating between an admitted and rejected state. One crucial aspect is how the source node decides which of multiple possible routes to use (recall that a source node may receive multiple route replies). To increase stability, we add a "on current path" flag in the RREQ and RREP packets, and force the source nodes to reuse a current path if available. Specifically, when the source node initiates route refreshing, it sets the "on current path" flag to 1. When an intermediate node receives RREQ and it is not on the current path it zeros the flag; otherwise, it leaves the flag unchanged. The destination copies the flag from the RREQ into the RREP packet. Upon receiving RREPs, the source node will select the path with flag 1 over those with 0.

5. PERFORMANCE ANALYSIS

To evaluate the performance of the admission control mechanism, we simulate several random and controlled wireless ad hoc network topologies. All simulations are conducted on the ns-2 network simulator with default parameters. Every node has a transmission range of 250m and carrier-sensing range of 550m. IEEE 802.11 MAC is used as the MAC layer with channel capacity of 2Mbps. We have implemented the admission control protocol using the Uppsala LUNAR implementation.

We measure quality of service in terms of high packet delivery rate with low variance. The accuracy of channel availability estimation mechanism can be judged by the number and size of flows admitted, while maintaining good quality of service.

5.1 QoS Enhancement Through Admission Control

We show through several small controlled experiments that the QoS routing and admission control results in high quality of service for all admitted flows, in situations where the lack of the QoS scheme leads to high packet loss and large delay.

In the first experiment, in the network shown in Fig. 4, flow 1 from node 9 to 16 starts at 10.0s, with bandwidth requirement of 380kbps, or link utilization requirement 0.19. 10 seconds later, node 11 starts a flow to node 18 and also requires bandwidth of 380kbps. At 30.0s, a third flow is started by node 17, attempting to reach node 10 and requiring 450kbps bandwidth. We observe the performance of each of the three flows in terms of throughput and

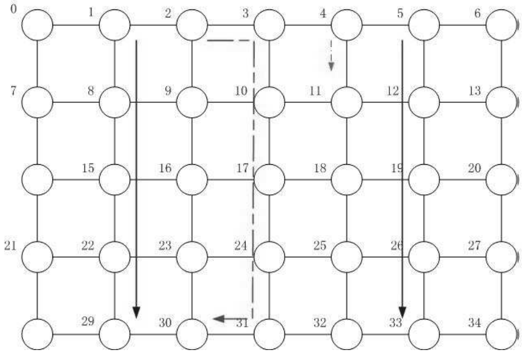


Figure 4: Simulation Network topology

average packet delay varying with time. The simulation results are presented in Fig. 5-6.

Fig. 5(a) shows the throughput of each of the 3 flows for basic LUNAR (without QoS support). As expected, the graph illustrates that as the simulation progresses and more sources become active, the channel becomes congested. As a result, the throughput of all the flows shows a significant instability. Fig. 6(a) presents the average delay of the received packets. Once the channel becomes congested, the delays for all the three flows increase greatly, along with large variation. Such a high delay experienced by the received packets is often unacceptable for real-time applications.

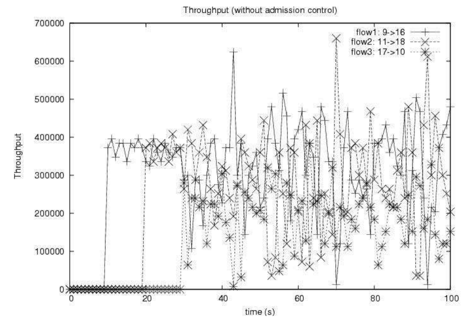
It is evident that in contrast to the poor performance without QoS support, the proposed QoS routing and admission control scheme enables admitted flows to experience much better service. To avoid congestion, the third flow is rejected due to lack of bandwidth. Compared to Fig. 5(a), Fig. 5(b) shows that traffic throughput for each of the admitted flows is nearly constant and matches their requirement. The delay, shown in Fig. 6(b), is extremely small. Note that the difference in the scale of the y-axis between Fig. 6(a) and 6(b) is two orders of magnitude. The short packet delay, consistent throughput and packet delivery fraction demonstrate that our scheme can be used to sustain real-time traffic applications, such as video and multimedia.

Next, we conduct a simulation in a $1000m \times 1000m$ static network with 20 randomly positioned nodes. Each node attempts to establish a CBR connection to a randomly chosen destination at random time. A sample run of the simulation is given in Fig.7

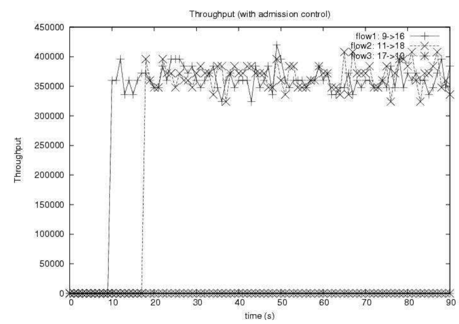
Each vertical line in the figure represents the start of a new flow, so we can see that only 7 flows have been admitted. However, the admitted flows achieved a constant throughput around 100kbps with low variance, matching their requirements. Repetitions of this randomized experiment showed similar results.

5.2 Benefit of Parallel Transmission Consideration

To analyze the benefit of the parallel transmission consideration, we consider a $1000m \times 1000m$ static network with 20 randomly positioned nodes. Every node in the network attempts to establish a CBR connection with a randomly chosen destination. All the links have the same bitrate 1.8Mbps. The packet size and the sending rate are randomly chosen between x and y. We compare the number of admitted flows and the aggregate network throughput when parallel transmissions are considered and not considered. The experiment has been run 53 times, each with different topology and traffic pattern. Fig 8(a) compares the number of admitted flows for each run of the experiment. Aggregate end-to-end

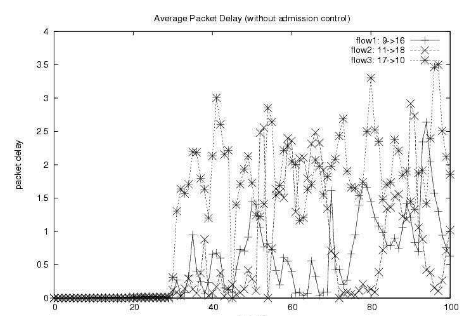


(a) throughput without admission control (dramatically changing, poor performance)

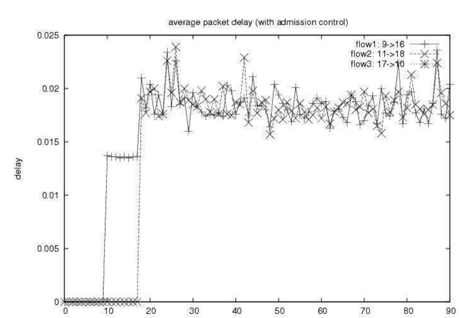


(b) throughput with admission control (stable as desired)

Figure 5: Admission Control vs. No Admission Control - Throughput



(a) average delay without admission control (large delay with huge variation)



(b) average delay with admission control (small delay with low variation)

Figure 6: Admission Control vs. No Admission Control - Average End-to-End Delay

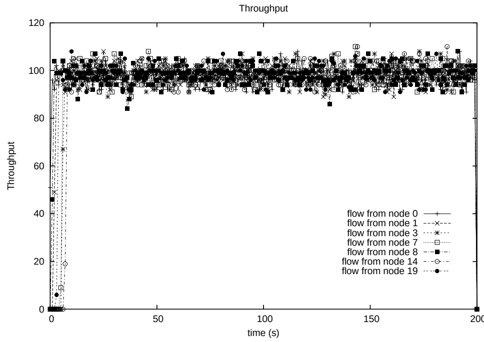
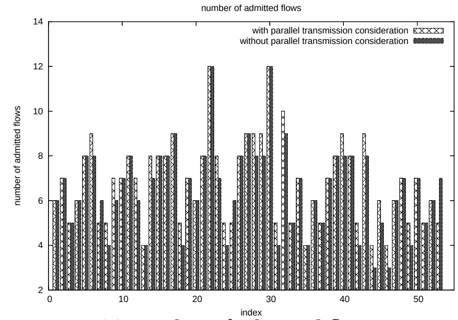


Figure 7: In the network with 20 randomly positioned nodes, only 7 (out of 9) flows are admitted; the 7 admitted flows achieve their desired throughput

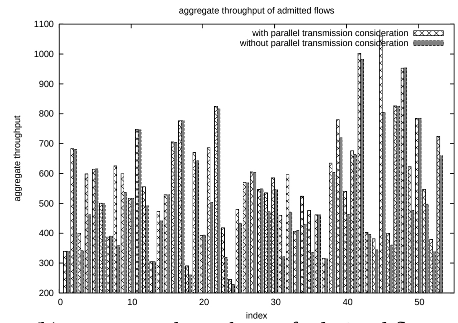
throughput is given in Fig 8(b). In these two figures, performance improvement can be seen from two aspects: 1) number of admitted flows, and 2) total volume of admitted flows. Out of 53 runs of the experiment, there are 19 times in which more flows are admitted with parallel transmission consideration. End-to-end throughput is also improved in these cases. There are in total 31 runs in which aggregate network throughput is higher with parallel transmission consideration. The biggest improvement occurs in the eighth run. When parallel transmission is considered, the aggregate end-to-end throughput is around 1.8 times of that without parallel transmission consideration. To understand this scenario, we illustrate the network topology and traffic pattern of the eighth run in Fig 9. Flow $1->14->13$ is the cause of the difference, i.e., it is admitted only when parallel transmission is considered. Since node 14 is the bottleneck of this flow, we only analyze its resource availability here. Node 14's carrier-sensing range is denoted as the dotted circle. After the first four flows were admitted, the flow $11->0$ and two hops of the flow $12->5$ occupy $48\times 2+22\times 2=140kbps$ bandwidth. The local utilization is thus $\rho_{local}=\frac{140}{1800}$, which is required by the transmissions within the whole *neighbor-carrier-sensing range*. Therefore, ρ_{local} is part of ρ_{csn}, and the other part of it comes from outside the *carrier-sensing range*, which we denote as ρ_{outcs}. The flows $5->15$, $18->3$ and one hop of the flow $11->5$ are outside node 14's *carrier-sensing range*. They made $\rho_{outcs}=(180\times 5+170\times 2+50\times 1)/1800=1290/1800$ and thus $\rho_{csn}=\rho_{local}+\rho_{outcs}=1430/1800$. For node 14, $\rho_{avail}=1-\rho_{csn}=370/1800$. If not considering parallel transmission, flow $1->13$ would get rejected at node 14, because ρ_{aggr} of the flow would be $285\times 2/1800=570/1800$, which is larger than ρ_{csn}. Otherwise, if parallel transmission is considered, channel availability is checked according to (6), which can admit the flow.

We also show that the estimation accounting for parallel transmission does not compromise the QoS of admitted flows while allowing more traffic into the network. In Fig 10, x-axis is the improvement in aggregate throughput with parallel transmission consideration; the two points corresponding to each value represent the packet delivery ratios with and without parallel transmission consideration in the corresponding experiment. In most experiments, parallel transmission admitted 50-150kbps more traffic without a noticeable reduction in packet delivery ratio. Thus, parallel transmission does not admit more traffic at the price of QoS. It brings more traffic into the network because it can better estimate available resources.

Fig 11 and Fig 12 show two case studies of end-to-end through-



(a) number of admitted flows



(b) aggregate throughput of admitted flows

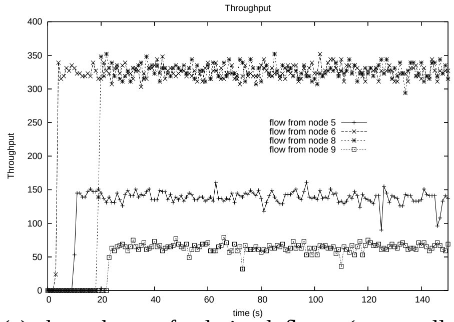
Figure 8: number of admitted flows and aggregate throughput of admitted flows. 19 experiments show that more flows can be admitted with parallel transmission consideration; including them, 31 experiments show the improvement in aggregate throughput

put over time. Each vertical line represents the start of a new flow. In the first case, the same number of flows are admitted, but the last admitted flow (from node 2) has larger volume when parallel transmission is considered, which leads to higher aggregate throughput. On the contrary, if parallel transmission is not considered, the flow from node 2 will be rejected at 21s and later on at 23s a smaller flow (from node 9) is admitted. In the second case, one more flow is admitted with parallel transmission consideration. This flow has end-to-end throughput of 160kbps, which means 160kbps increase in the aggregate end-to-end throughput. Even though more traffic has been admitted, there is no evident increase throughput variance.

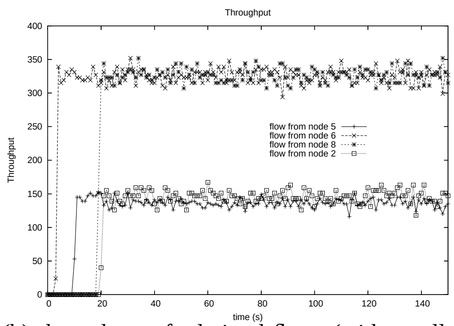
6. CONCLUSIONS

We have presented an admission control mechanism integrated with ad hoc routing for 802.11 wireless ad hoc networks. It performs bandwidth aware routing which discovers the route satisfying the flow bandwidth requirements and admission control which determines whether a flow can be admitted. The admission control algorithm is able to calculate available bandwidth and predict the bandwidth consumption of a flow while taking into account parallel transmissions. Simulation results show that integrating admission control into the routing protocol enables the identification of alternate routes if the shortest path is congested. Admission control can limit the amount of data traffic in the network to provide QoS guarantees to admitted flows. By exploiting channel reuse, our scheme can admit more traffic while maintaining QoS compared to mechanisms that do not account for parallel transmissions.

7. REFERENCES

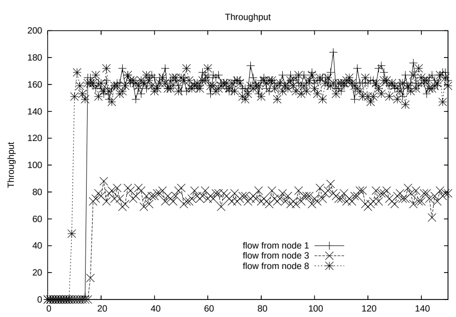


(a) throughput of admitted flows (no parallel transmission consideration)

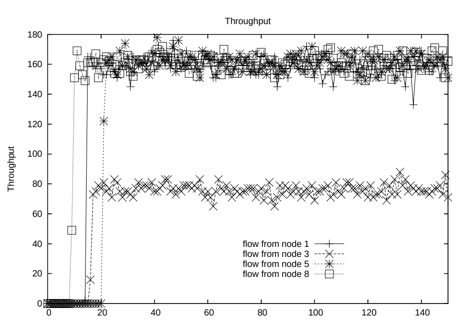


(b) throughput of admitted flows (with parallel transmission consideration)

Figure 11: Case Study 1 - Benefit of Parallel Transmission Consideration. The same number of flows are admitted, but not considering parallel transmission leads to rejecting a bigger flow while admitting a smaller one started later



(a) throughput of admitted flows (no parallel transmission consideration)



(b) throughput of admitted flows (with parallel transmission consideration)

Figure 12: Case Study 2 - Benefit of Parallel Transmission Consideration. One more flow is admitted with parallel transmission consideration

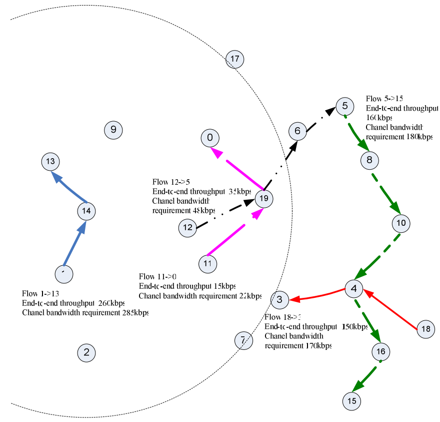


Figure 9: Example of parallel transmissions, only when considering overlap due to parallel transmissions, flow $1-14-13$ can be admitted

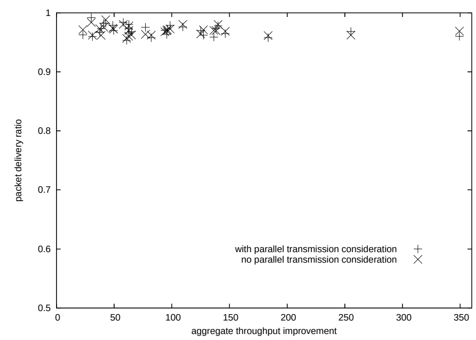


Figure 10: Improvement in aggregate throughput with parallel transmission consideration; in most experiments, parallel transmission can admit $50-150kbps$ more traffic without a noticeable reduction in PDR.

[1] IEEE Computer Society, "802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications"

[2] L. L. Peterson and B. S. Davie, *Computer Networks: A System Approach*, Morgan Kaufmann Publisher, 2nd edition, Oct 1999.

[3] K. Wu and J. Harms, "QoS Support in Mobile Ad Hoc Networks,"*Crossing Boundary - an interdisciplinary journal*, VOL1, No 1 - Fall 2001.

[4] Y.Dong, T. Yang and D. Makrakis, "SRL-Enabled QoS Model for Mobile Ad Hoc Networks," *International Conference of Communications, Circuits and Systems & West Sino-Exposition 2002(ICCCAS&WESINO EXPO'02)*. pp.414-418, June 2002.

[5] G. Ahn, A. T. Campbell, A. Veres, and L. Sun, "SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks," *Proc. IEEE INFOCOM'2002*, New York, June 2002.

[6] Y. Dong, D. Makrakis and T. Sullivan, "Effective Admission Control in Multiple Mobile Ad Hoc Networks," *IEEE Proceedings of ICCT2003*, 2003.

[7] M. G. Barry, A. T. Campbell, and A. Veres, "Distributed Control Algorithm for Service Differetiation in Wireless Packet Networks," emphProc, IEEE INFOCOM'2001, New York, April 2001.

[8] C.R. Lin and C. Liu. "An On-demand QoS Routing Protocol for Mobile Ad Hoc Networks," *IEEE Global Telecommunications Conference (GLOBECOM)*, 2000.

[9] I. D. Chakeres and E. M. Belding-Royer. "PAC: Perceptive Admission Control for MobileWireless Networks," *Proceedings of the 1st*

International Conference on Quality of Service in Heterogeneous Wired/Wireless Network (QShine)*, Dallas, TX, October 2004.

[10] Y. Yang and R. Kravets. "Contention-Aware Admission Control for Ad Hoc Networks," *Technical Report 2003-2337*, University of Illinois at Urbana- Champaign, April 2003.

[11] C. Tschudin and R. Gold. "LUNAR - Lightweight Underlay Network Ad-Hoc Routing," *tech. rep.*, University of Basel, Switzerland, January 2002.

[12] C. Tschudin and R. Gold. "SelNet: A Translating Underlay Network," *http://www.docs.uu.se/docs/research/projects/selnet*

[13] L. Luo, H. Liu and D. Raychaudhuri. "A QoS Routing and Admission Control Scheme for Wireless Ad-Hoc Networks," *Technical Report*, WINLAB, Rutgers University, June 2005.

[14] K. Fall and K. Varadhan. "The ns manual," *The VINT Project*, UC Berkeley, LBL, USC/ISI, Xerox PARC, 1997. Available at: http://www.isi.edu/nsnam/ns/.

[15] The CMU monarch project, "Wireless and mobility extensions to ns-2," *http://monarch.cs.cmu.edu/cmu-ns.html*.

[16] D. Maltz. "Resource management in multi-hop ad hoc networks," *Technical Report CMU CS 00-150*, School of Computer Science, Carnegie Mellon University, July 2000.

[17] R. S. Prasad, M. Murray, C. Dovrolis, and K. Claffy, "Bandwidth Estimation: Metrics, Measurement Techniques and Tools," *IEEE Network*, 17(6), November 2003.

[18] G. Chakrabarti and S. Kulkarni, "Load Balancing and Resource Reservation in Ad-hoc Networks," *Ad-Hoc Networks*, 2004.