A Distributed QoS MAC Protocol for Wireless Mesh

Mathilde Benveniste
Avaya Labs - Research
benveniste@ieee.org

Abstract

Abundant hidden node collisions and correlated channel access due to multi-hop flows degrade QoS in wireless mesh networks. QoS in nearby WLANs operating on a single channel is also affected. We propose using wider contention windows for backoff to lower the risk of repeated hidden-node collisions, a spatial extension of the TXOP concept called 'express forwarding' to clear multi-hop flows sooner, and a new mechanism called 'express retransmission' to reduce collisions on retransmission. Simulation results show the potential benefit of the proposed enhancements and impact on fairness.

1. Introduction

A wireless mesh network is a network that accommodates forwarding of packet traffic on a wireless medium over one or more hops. A mesh may furnish wireless connections either to access points (APs) serving different WLANs, or simply to devices supporting peer-to-peer wireless communication. A wireless mesh shares many of the challenges encountered in mobile and ad hoc networks, also known as MANETs [1] – [3]. The IEEE 802.11s Task Group is currently addressing the standardization of a wireless mesh MAC that will be compatible with the IEEE 802.11 WLAN MAC protocol [4].

Examples of mesh usage include emergency early response, public Internet access, metropolitan hotspot coverage, and enterprise and campus wireless networks. A mesh network may be connected to the wired network through a portal. The number of channels used by a wireless mesh varies. Meshes providing wireless backhaul to a collection of APs typically require multiple channels.

QoS objectives can be pursued on different ISO layers. Routing may employ end-to-end latency as the optimization criterion [5]. While routes may adapt to topology changes and traffic, they tend to be static. Cross-layer interactions and their implications for QoS have also been considered [6].

MAC protocol design is important in meeting QoS requirements since much of the latency experienced in a wireless network occurs in accessing the shared medium. In addition, MAC protocols must be interoperable with existing wireless networks operating on the same RF spectrum and fair toward all users.

Latency restrictions for QoS are meaningful end-to-end. Of the 150 milli-sec limit recommended by ITU G.114, a budget of 40-50 milli-sec remains after accounting for delays caused by functions like encoding, packetization, decoding and jitter buffering [7]. If one extrapolates from experience with WLANs, meeting the above latency limit would not appear difficult for any but the longest multi-hop flows. We find, however, that wireless meshes have novel collision behavior that imposes latency increases on both mesh and co-channel WLANs beyond what non-mesh experience suggests.

1.1 MAC protocol design for QoS

MAC protocol design is an important aspect of meeting QoS, as a key contributor to latency in a wireless network is the contention occurring when accessing the shared medium. Access can be combined with channel assignment for meshes using multiple channels. If the number of transceivers on a node is smaller than the number of channels employed in the mesh, access can be combined with scheduling radio and channel use on different links [8] – [15].

Several MAC protocols exist for both single-channel and multi-channel meshes. For single-channel meshes, the IEEE 802.11 distributed MAC protocol for WLANs, known as EDCA [4], [16], is the protocol most commonly used. EDCA enables WLANs to meet QoS requirements through the TCMA (Tiered Contention Multiple Access) protocol for prioritized channel access [16], [17]. In the absence of low priority traffic, however, prioritized access does not offer any benefit. Consequently, EDCA results in comparable latencies to the basic CSMA/CA protocol [18], [19]. The following
questions thus arise: Considering distributed MAC protocols that are compatible with WLANs operating on the same channel as the mesh, does CSMA/CA provide the best QoS performance for a wireless mesh, or can another MAC protocol perform better?

The single-channel mesh is of special concern, both because of its potential value and because of the special challenges it presents. Though not appropriate for backhaul of multiple fully loaded WLAN access points, single-channel meshes will provide the technology toward which future WLANs will evolve. They can be used as a means of extending the range of an infrastructure wireless network and for data rate improvement. By replacing the WLAN access point with a mesh portal as the distribution network interface, wireless devices will be able to reach the wired network from a longer distance away, on multiple hops. Shorter multiple-hop transmission will also increase the realizable data rate relative to the rate achieved with a single hop.

Although channel assignment and radio scheduling problems do not arise in single-channel meshes, MAC design is more challenging. The short channel re-use distances encountered in single-channel meshes cause a prevalence of hidden nodes.

Before exploring how hidden nodes impact QoS performance of single-channel meshes, we describe in Section 2 the distributed MAC protocol for IEEE 802.11 WLANs and the remedy for hidden node collisions in WLANs. Section 3 describes how the mesh topology impacts the effectiveness of the 802.11 MAC protocol. A new MAC protocol and other remedies for removing the deleterious effects introduced by mesh topology are described in Section 4. In Section 5, we compare the performance of different MAC protocol options for static routing conditions. Section 6 contains conclusions.

2. The IEEE 802.11 MAC protocol

The IEEE 802.11 standard for WLANs employs a distributed MAC protocol, CSMA/CA [4], [16]. This protocol avoids collisions through carrier sensing, backoff, and handshake. A device transmits only when the channel is determined idle. Each device listens to the channel and, if it hears signal, postpones transmission and enters into the ‘backoff procedure’. Transmission is deferred by a time that depends on a random backoff value. Backoff facilitates collision avoidance between multiple stations that would otherwise attempt to transmit immediately after completion of the current transmission. The backoff value expresses, in time slots, the cumulative time the channel must be idle before access may be attempted.

IEEE 802.11 WLANs use TCMA, an enhanced version of CSMA/CA, to prioritize access among different traffic types [17]. A combination of prioritized access and admission control offer satisfactory QoS in IEEE 802.11 WLANs. Fairness among devices with one or multiple types of traffic is ensured through the use of different EDCA queues for different types of traffic, each queue contending independently.

With CSMA/CA, a station engaged in backoff countdown must wait while the channel is idle for time interval equal to DIFS before decrementing its backoff delay immediately following a busy period, or before attempting transmission. According to the TCMA protocol, variable lengths of this time interval, which is called Arbitration-Time Inter-Frame Space (AIFS), lead to varying degree of accessibility to the channel. A shorter AIFS will give a higher-priority traffic queue an advantage in contending for channel access. Use of TCMA meets packet latency requirements in a WLAN when reasonably loaded. The challenge is to meet similar per flow end-to-end requirements for a reasonably loaded mesh.

3. Using the IEEE 802.11 MAC in mesh

Prioritized access increases the probability of high-priority traffic transmitting before lower priority traffic. However, that alone is not sufficient to meet the latency restrictions for QoS. The end-to-end delay experienced in a mesh multi-hop path may be longer than a simple multiple of the delay experienced for a single hop in a non-mesh environment. The prevalence of hidden nodes and the interaction of contention-based access with multi-hop flows increases collision rates and retransmissions, and leads to higher channel utilization per attempted transmission and ultimately to dropped frames and/or latency increases.

3.1. Hidden node collisions

Collisions in wireless mesh networks occur for two reasons. One type of collision occurs if the backoff delay of two of more such devices waiting to transmit expires simultaneously. Another way collisions arise is from ‘hidden nodes’ [20]. A hidden node is one that cannot sense an ongoing transmission, but if it transmits, it can interfere with the decoding of such transmission at the receiver. An
example is illustrated in Figure 1, where node B is unable to decode a transmission from A if it is within interference range of node F, which transmits simultaneously because it cannot hear or be heard by node A. This is known as a ‘hidden node collision’.

Figure 1. Hidden node collision

The IEEE 802.11 MAC protocol offers a remedy for hidden node collisions. Short control frames – namely, RTS and CTS frames [21] – by the sender and the receiver. The RTS is sent by the source of the pending transmission. The frame includes the period of time for which the channel is reserved. The receiver returns a CTS control frame if the channel is clear to send. This frame also notifies the neighboring nodes of the channel reservation as it carries a field with the duration of the channel reservation. The RTS/CTS handshake protects against hidden node collisions in two ways. If the receiver of the RTS is hearing another transmission undetected by the sender of the RTS, the CTS will not be sent. This preempts a hidden node collision involving the frame. Once the frame transmission starts, any hidden nodes will refrain from transmission because they received the CTS, thus averting hidden node collisions.

Use of RTS/CTS has the penalty of increased bandwidth taken by the control frames. Additionally, collisions are not entirely avoided, as both the RTS and CTS may be involved in collisions. Thus, RTS/CTS is most useful typically for protecting long frames on slow channels. If the header of a frame is a significant portion of a frame’s transmission time, as for instance with IEEE 802.11a/g/n channels (due among others to the lower transmit rate for the header), the benefit of the handshake is outweighed by the added overhead of the control frames, which reduces available bandwidth, and increases collision rates and end-to-end delays [22].

Hidden node collisions arise on uplink transmissions in WLANs, or in WLANs with overlapping coverage areas. The latter can avoid cross collisions by selecting different channels. Hidden node collisions are more prevalent in single-channel mesh networks than in WLANs. Hidden nodes arise also when a mesh is located near a WLAN that uses the same channel. Hidden nodes are especially prevalent in mesh networks used for range extension, as a node cannot sense the transmissions by neighbors of neighbor nodes. Between a pair of potentially non-interfering nodes in a connected mesh, lies a third node that can communicate with both pair members, operating on the same channel.

The prevalence of hidden nodes increases collision rates and retransmissions, leading to higher channel utilization per attempted transmission and to dropped frames. In the example of Figure 1, nodes A and F cannot hear one another while node B and E can hear both A and F. The transmissions A and F to B and E, respectively, overlap in time. As a consequence, both B and E experience collisions. The backoff delay of each is decremented in time even when the other is transmitting, and transmission is likely to be attempted while the other is transmitting. Repeated collisions increase latency. If the retry limit is reached, their frames are dropped. With adjustable data rates, high dropped-frame rates lead to data rate reduction and low throughput.

3.2. Multi-hop flows

In addition to repeated hidden node collisions along its hops, the interaction of multi-hop flows with contention-based access can cause latency increases on other flows. This novel behavior of meshes can impact nearby WLANs as well.

Latency increases can be caused by multi-hop flows to a nearby transmission that is involved in a collision. According to the IEEE 802.11 MAC protocol, a device attempting a failed transmission must draw a random backoff from a wider range – known as the contention window. Because of its longer retry backoff, the retransmitting device is at a disadvantage relative to transmissions of a multi-hop flow that are attempted for the first time, and may have to wait for the completion of multiple hops of that flow before retransmission is possible. Transmissions near a multi-hop path are especially vulnerable, as collisions are often caused by acknowledgements along a multi-hop flow as it advances along its path. Figure 1 illustrates how a multi-hop flow may delay a transmission near its path and make it wait for the entire multi-hop flow to
complete. The acknowledgement from node B to node A causes a collision for the transmission to node E. Because of its longer backoff, node F will probably have to wait for nodes B and C to forward the frame they receive before it can retransmit.

Applications with short frame inter-arrival times (e.g. HDTV) risk going unstable if situated near multi-hop flows. The sooner the multi-hop flow completes the sooner retransmission will succeed.

4. MAC remedies for wireless mesh

We propose three measures to improve the QoS performance of single-channel wireless meshes. They are: (1) use of wider contention windows for transmission retry following a collision, (2) ‘express forwarding’ and (3) ‘express retransmission’.

4.1. Wide Retry Contention Window

By increasing the contention window on transmission retry, the likelihood of averting a repeat collision increases for two mutually hidden nodes. Since the backoff delay in this case represents the clock time − not the cumulative channel idle time − each such node will wait before transmitting, increasing the retry contention window decreases the probability that the transmissions of the two nodes will overlap in time.

This measure can be implemented simply when using the IEEE 802.11 MAC protocol by raising CWmax, the contention window size such that once it is reached, the contention window size is no longer doubled after a collision.

4.2. Express Forwarding

‘Express forwarding’ is an enhancement of the CSMA/CA protocol designed to reduce the latency experienced end-to-end by a multi-hop wireless mesh [23]. Multi-hop transmissions are expedited by having the transmitted frame reserve the channel on each leg of the multi-hop path for the next hop. When a frame is express forwarded, the channel is reserved by extending the Duration field value of the frame enough to silence all neighboring nodes and give the receiving node the opportunity to seize the channel and forward the frame, as illustrated in Figure 2. The acknowledgement to a data frame or the CTS sent in response to an RTS frame contains a Duration value derived from the value in the frame for which it is returned, adjusted for elapsed time. This way, following the contention for the transmission on the first hop, an express-forwarded frame is transmitted quickly from hop to hop without contention, causing multi-hop end-to-end delays to decrease.

The notion of an Express Forwarding TXOP (EF-TXOP) thus arises, which is a time-space extension of the IEEE 802.11 TXOP [4], [16]. Transmit opportunities (TXOPs) enable a source to transmit multiple frames following a single successful channel access attempt, without having to contend for the channel. In an EF-TXOP, consecutive linked transmissions of a multi-hop flow are made without the need to contend more than once. In a TXOP, the right to transmit contention-free following the initial successful channel access attempt remains with the source of the transmission. With the EF-TXOP, the right to access the channel contention-free is handed over to the next node along the multi-hop path.

Figure 2. Express Forwarding reservation

Because all but one frame in an EF-TXOP is transmitted without contention, EF-TXOPs help reduce the frequency of collisions, as done by TXOPs. This increases channel use efficiency. As in the case of TXOPs, a limit can be imposed on the maximum length of an EF-TXOP, in order to avoid excessive delay increase for non-express-forwarded traffic. Because express forwarding uses carrier sense functions and backoff collision avoidance, it can interoperate with WLANs using the same channel.

The time interval added to the Duration field to reserve the channel for express forwarding should be one time slot plus the shortest time necessary to ensure that IP processing of the transmitted frame is complete at the receiving node. The additional reservation time gives the forwarding node the opportunity to seize the channel before any of its neighbors, as they have set their NAV timer according to the received frame Duration field value. If processing of an incoming frame commences as
soon as it is received, and in parallel with the 
acknowledgement, the time increment added to the 
Duration field is the time by which the processing 
time exceeds the time it takes to send an 
acknowledgement, if any, plus one time slot. The 
Duration field value of an express-forwarded frame 
is not extended on the last hop of a multi-hop 
transmission.

EF-TXOPs can be combined with TXOPs in 
several ways. An express-forwarded frame can be 
transmitted along a hop as part of a TXOP. In order 
to enable the receiving node to seize the channel 
without contention for the next hop, the channel must 
be reserved beyond the end of the TXOPs 
transmission and acknowledgement.

When the received express-forwarded frame must 
be forwarded, if other frames queued at the receiving 
node can be sent in the same TXOP, the entire TXOP 
may go contention free. Its transmission may start 
immediately after the receiving node sends the last 
acknowledgement and following the appropriate 
AIFS idle period, even though the backoff delay of 
the frames included in this TXOP may not have 
expired. Transmission may thus start before the 
received express-forwarded frame is fully processed.

The RTS/CTS handshake is unlikely to benefit 
performance of wireless networks operating on fast 
channels, like IEEE 802.11a/g/n, for the reasons 
given earlier. For slower channels, the handshake 
can improve performance. RTS/CTS helps in a 
different way than express forwarding. Express 
forwarding can be used to send an RTS along each of 
the legs of a multi-hop path. This way RTS/CTS can 
reduce the penalty from forward hidden node 
collisions, while express forwarding will expedite the 
multi-hop flow and reduce the contention 
experienced by the RTS along the multi-hop path.

### 4.3. Express Retransmission

Retransmission of a failed transmission typically 
involves drawing a backoff delay from a wider 
contention window than the initial transmission 
attempt. An expedited retransmission, referred to as 
‘express-retransmission’, can be sent contention free 
if the source retransmits as soon as the 
acknowledgment timer expires. If collision is 
experienced for an express-retransmitted frame, 
further attempts to transmit this frame will involve 
backoff from a widened contention window.

Express retransmission of express-forwarded 
frames helps shorten the end-to-end latency of a 
multi-hop flow. An express-retransmitted frame will 
not collide with transmissions from neighbors as they 
have their NAV still set according to the duration 
field of the express-forwarded frame. If the collision 
that prompted the retransmission was due to a hidden 
node, the collision is less likely to repeat than in the 
case where both re-transmissions are attempted with 
backoff. The express retransmission occurs without 
backoff, while other retransmissions must use a long 
backoff delay. An exception occurs if the hidden 
node collision involves another express-forwarded 
frame. Collision is likely on the first retransmission 
attempt, but less likely on the subsequent attempt, 
since the backoff procedure is invoked with 
contention windows widened by a factor of four.

### 5. Performance evaluation

The performance benefits of express forwarding 
and express retransmission have been demonstrated 
in several studies for a range of scenarios [24], [25].

#### 5.1. Description of study

The objective of these studies was to compare the 
QoS performance of a lightly loaded mesh, co- 
located with WLANs using the same channel for 
diverse access scenarios. We present here 
results from a study that deals with three scenarios, 
as described in Table 1. In the first scenario, all 
traffic accesses the channel through the IEEE 802.11 
EDCA mechanism. In the second, express 
forwarding is employed for the multi-hop flows use. 
In the third scenario, the multi-hop flows use express 
forwarding and express retransmission.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. EF Disabled</td>
<td>Express Forwarding disabled</td>
</tr>
<tr>
<td>2. EF Enabled</td>
<td>Express Forwarding enabled for multi-hop flows</td>
</tr>
<tr>
<td>3. EF-ERTX Enabled</td>
<td>Express Forwarding &amp; Express Retransmission enabled for multi-hop flows</td>
</tr>
</tbody>
</table>

The network configuration consists of three 
WLANs and a wireless mesh, all operating on the 
same channel. The network traffic consists of 
constant flows between specified end points. The 
traffic flows simulated are three 3-hop flows, and a 
collection of single-hop flows. The multi-hop flows, 
which are part of the mesh, carry VoIP calls outside 
the mesh through a gateway device, the mesh portal. 
The single-hop flows belong either to the WLANs or
to the mesh. The traffic of these flows is VoIP, low-resolution video, or high-resolution video, as indicated in Figure 3. The IP phones generate bidirectional streams communicating either with mesh peers or with the outside world through Node 0, which is the mesh portal. There was no node mobility; static routing is employed. Table 2 presents the key traffic and MAC parameters.

All nodes were equipped with a single 802.11a radio. The channel was assumed to be noise-free. Application data traffic was transmitted at 52 Mbps and acknowledgments at 24 Mbps. A 50 μsec IP processing delay was assumed at each node.

### 5.2. Results

Table 3 presents the mean end-to-end delays for all the flows under the three scenarios described in Section 5.1. The table indicates the network to which each flow belongs and whether it is a multi-hop flow – marked as (M) – or a single-hop flow – marked as (S). Figures 4 and 5 present, respectively, the normalized number of retransmissions and dropped frames by transmitting node. Normalization was done by dividing by the number of frames for which a transmission attempt was made at a given node.

![Figure 3. Network layout](image)

Simulations were conducted by using the OPNET Modeler modeling platform [26].

### Table 2. Key traffic and MAC parameters

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Payload (Bytes)</th>
<th>Frame Spacing (μsec)</th>
<th>CWmin+1</th>
<th>CWmax+1</th>
<th>WiMAX/Mesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP call</td>
<td>200</td>
<td>20</td>
<td>7</td>
<td>15/1023</td>
<td>31/1023</td>
</tr>
<tr>
<td>Low-resolution Video</td>
<td>1464</td>
<td>20</td>
<td>283</td>
<td>15/1023</td>
<td>31/1023</td>
</tr>
</tbody>
</table>

*CWmin+1 is the contention window size used to draw a backoff delay when a transmission is first attempted

**CWmax+1 is the maximum size the contention window may assume when retransmission is attempted following a collision

![Figure 4. Normalized retransmissions by node](image)

Of the three multi-hop flows, only one – the call to Node 3 – meets the latency requirements for QoS when EDCA is the access mechanism. The other two multi-hop flows experience excessive delays and retransmissions. On some nodes, the average number of attempts needed exceeds two per frame. Retransmissions cause frames to be dropped; as many as 4 per cent of the frames are dropped at Node 11.
When express forwarding is applied to the multi-hop flows, the latency on all flows is reduced, whether they are express-forwarded or not. The latency reduction is greater for the flows that are express forwarded, but the other flows benefit as well. The number of retransmissions declines and the number of dropped frames is halved. All calls can meet QoS requirements with express forwarding. Express retransmission, combined with express forwarding, further improves MAC performance.

Figure 5. Normalized dropped frames by node

These results, as well as the other performance studies cited here, suggest that when packets are transmitted on a reserved channel, rather than contend for the channel on every leg of a multi-hop path, total contention is reduced considerably. As a consequence, both multi-hop and single-hop flows benefit from use of express forwarding for the multi-hop flows.

6. Summary and Conclusions

This paper deals with meshes using a single channel for all mesh nodes, and a single radio per mesh node. It describes a novel MAC protocol for mesh, called Express Forwarding, which represents an enhancement of the CSMA/CA protocol. Express Forwarding can be further enhanced through Express Retransmission. Express Forwarding can coexist with WLANS using the standard IEEE 802.11 MAC protocols to access the same channel as the mesh.

The performance of the new protocol was examined for a single channel mesh that is co-channel with several nearby WLANs. The combined traffic load was similar to that seen in a WLAN, and the multi-hop paths were of moderate length. It was observed that express forwarding was able to deliver delay performance that meets the QoS requirements for real-time applications, while the standard IEEE 802.11 EDCA access mechanism could not.

Simulations confirmed that both types of frames (express-forwarded frames and non-express forwarded frames) enjoy shorter latencies when express forwarding is used for multi-hop transmissions. Paradoxical as this may seem, giving preferential treatment with express forwarding to nodes forwarding multi-hop traffic over nodes that transmit traffic for a single hop, has helped both types of transmissions. This is because, as with a TXOP, the EF-TXOP reduces contention on the channel and thus decreases the collision probability. Fewer collisions imply shorter latencies for all traffic. As in the case of TXOPs (where single-frame latencies may increase as a result of TXOP use), there may be some non-express forwarded traffic whose short delays will increase somewhat. According to our simulations, such increases are small and the resulting single-hop latencies are far shorter than the multi-hop latencies. As an added precaution, however, one can impose a limit on the maximum length of an EF-TXOP, very much the way we limited the maximum length of a TXOP.

The simulation studies involved VoIP and video traffic only, for the transmission of which the channel is accessed with the same AIFS. No Best Effort (lower priority) traffic was included. Had lower priority traffic been included, EDCA would have prioritized access accordingly. Express forwarding and prioritized access are orthogonal mechanisms that can be used together.

Express forwarding is a fair MAC protocol. When analyzing fairness in channel access on a per-node basis, express forwarding gives preferential treatment to nodes forwarding multi-hop traffic over nodes that transmit traffic for a single hop. Since the user’s experience is tied to the end-to-end latency, however, fairness should be considered on a per-flow basis. Express forwarding is fairer than EDCA as it helps reduce multi-hop flow latencies and prevents single hop flows from experiencing longer delays than multi-hop ones. It is important to note that the traffic disadvantaged with express forwarding – namely, the single-hop traffic – enjoys better performance when express forwarding is employed than when it is not. In general, all traffic enjoys better QoS performance with express forwarding than with EDCA.

Express forwarding can be extended to apply to multi-channel meshes. This will be the subject of future investigation.
References


[22] M. Benveniste and K. Sinkar, “Performance comparison of RTS/CTS and ‘Express Forwarding’ for mesh”, IEEE Doc 802.11-08-0291r0


