QoS for Wireless Mesh: MAC Layer Enhancements[*]

Mathilde Benveniste, Ph.D.
En-aerion
benveniste@en-aerion.com

Abstract

Wireless mesh networks present MAC design challenges beyond those of WLANs. Abundant hidden nodes increase the number of collisions. This, combined with the correlated access needed when forwarding a multi-hop flow, degrades QoS. MAC enhancements for meshes are presented in this paper that reduce latency for mesh traffic while promoting co-existence with nearby WLANs. Wider contention windows for backoff lower the risk of repeated hidden-node collisions, a spatial extension of the TXOP concept called 'express forwarding' clears multi-hop flows sooner, and a new mechanism called 'express retransmission' reduces collisions on retransmission. Simulation results show the potential benefit of the proposed enhancements. The issue of fairness is addressed, as well as preservation of QoS in nearby WLANs.

1. Introduction

A wireless mesh network is a network that accommodates forwarding of packet traffic on a wireless medium over one or more hops. Enabling multiple-hop communication, which gives rise to a mesh network, extends the range of a wireless LAN (WLAN). 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 gateway, the *portal*, facilitates communication of the users of the mesh with users on other networks. A wireless mesh shares many of the challenges encountered in mobile and *ad hoc* networks, also known as MANETs [2] – [4].

Wireless mesh is useful both in environments where wired network infrastructure is unavailable and where eventual connectivity to the available wired network is desirable. The first, commonly known as *ad hoc* mode meshes, are useful for the ability to be quickly deployed with low cost where there is no wired infrastructure. The second type of mesh, known as *infrastructure* mode meshes, help extend connectivity range without additional wiring. Examples of mesh usage include emergency early response, public Internet access, metropolitan hotspot coverage, and enterprise and campus wireless networks.

MAC design for wireless meshes must account for a variety of features, such as the number of physical channels used in the mesh. Low volume meshes linking devices through peer-to-peer single- or multi-hop wireless connections can perform well on a single channel. Meshes providing wireless backhaul for a collection of APs, however, would require greater channel capacity than any one of those APs. Multiple channels would be needed to backhaul traffic of multiple APs. Finally, different MAC protocols are needed when multiple channels are used in a mesh with a mix of multiple radios per device. 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 [5].

1.1 QoS in Wireless Networks

QoS objectives can be pursued on different ISO layers. QoS metrics such as end-to-end latency can serve as the optimization criterion in routing. Routes may vary in time due to mobility and topology changes [6] - [8]. Routes can also be adapted to traffic-trend changes over time, but routes do not change on a per-packet basis. The excessive control load needed to change routes would defeat an attempt to use routing to resolve collisions. Cross-layer interactions and their implications for QoS have also been considered [9] - [11].

Since much of the latency experienced in a wireless network occurs in accessing the shared

---

[*] Some of this work was done while the author was affiliated with Avaya Labs – Research.

medium, MAC protocol design is important in meeting QoS requirements. Whether transporting packets for backhaul to/from APs or linking wireless devices over multiple hops, applications with limited latency tolerance should be delivered within the required delay bounds. In addition, MAC protocols must be compatible with existing wireless networks operating on the same RF spectrum. Interoperability implies fair behavior toward other users of the RF spectrum, and especially not destroying the QoS expected by such users. The underlying problem is that of accessing the wireless medium in a fair, efficient, and distributed manner.

Latency restrictions for QoS are meaningful end-to-end. International Telecommunications Union document G.114 recommends a limit for end-to-end delay of 150 milli-seconds for real-time voice [12]. After subtracting from this total delay budget 50 to 60 milli-seconds for encoding, packetization, decoding and jitter buffering delays, the delay allowed for a wireless mesh carrying real-time traffic will depend on other delays experienced outside the wireless mesh. If the wireless mesh stands alone it will have a greater delay budget than if it interfaces with other network infrastructure. Voice over IP packets traversing wired networks experience IP network delays of about 50 milliseconds, which include propagation, table lookup and queuing delays. The delay budget would typically leave voice traffic between 40 and 50 milliseconds for network access/egress. The mesh latency limit applies on a per flow basis. Hence, in a wireless mesh, the allowed delay restriction applies to the entire multi-hop path.

If one extrapolated from experience with WLANs, meeting the above latency limit would not appear difficult for any but the longest multi-hop flows. If latency for single-hop access was less than 10 milliseconds, a five-hop path could be completed within the allowed time limit. 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.

1.2 MAC protocol design for real-time traffic over mesh

A key contributor to latency in a wireless network is the contention occurring when accessing the shared medium. Hence, the design of the MAC protocol is an important consideration in meeting the requirements of real-time traffic. The mutual RF interference experienced at nodes sharing the same channel, which is added to ambient noise, can prevent correct decoding on the receiving node.

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, [5], [13] is the MAC protocol most commonly used [4]. For meshes using multiple channels, access can be combined with channel assignment. In addition, 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 [14] – [23]. Of the multi-channel protocols, some employ EDCA as the underlying MAC protocol and interoperate with the IEEE 802.11 MAC and some do not.

EDCA enables WLANs to meet QoS requirements through the TCMA (Tiered Contention Multiple Access) protocol for prioritized channel access [13], [24]. In the absence of low priority traffic, however, prioritized access does not offer any benefit. Consequently, EDCA results in comparable latencies with the basic CSMA/CA protocol [25], [26]. The following question thus arises:

Considering distributed MAC protocols that are compatible with WLANs operating on the same channel as the mesh, does the CSMA/CA MAC provide the best QoS performance for a wireless mesh, or can another MAC protocol perform better?

The single-channel mesh is of special concern, for a variety of reasons. Single-channel meshes are expected to gain acceptance rapidly once standardized, and through the flexibility they offer, will provide the technology toward which future WLANs will evolve. Though not appropriate for backhaul of multiple fully loaded WLAN APs, 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 AP 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. Multi-hop transmission will also increase the realizable data rate. Devices situated on the edge of a WLAN’s coverage area are limited to transmit on a single hop at low data rates. With one or more devices situated in between, the edge device’s traffic would be forwarded on multiple yet shorter hops, which would be capable of higher rates.

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. The prevalence of hidden nodes increases collision rates and retransmissions, and leads to higher channel utilization per attempted transmission and ultimately to dropped frames. In the rest of the paper, we are concerned with meshes employing a single channel throughout the mesh and a single radio per mesh node. This paper is based on the author's presentation on this subject at MESH 2008 [1].

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 2 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 existing IEEE 802.11 MAC protocol

The IEEE 802.11 standard for WLANs employs a distributed MAC protocol, CSMA/CA. A combination of prioritized access and admission control offer satisfactory QoS in IEEE 802.11 WLANs. Prioritized access is achieved through service differentiation. Higher priority packets have a higher probability of accessing the channel before lower priority frames. 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 [5], [13].

The CSMA/CA protocol has been designed to avoid 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 busy, postpones transmission and enters into the 'backoff procedure'. This involves deferring transmission by a random time, determined by the backoff value drawn randomly. 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 [24]. 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 station an advantage in contending for channel access. Differentiation between different access categories is achieved by assigning a shorter AIFS to a higher priority access category.

Prioritized distributed channel access mechanisms like TCMA meet packet latency requirements when the WLAN is reasonably loaded. The challenge is to meet similar end-to-end QoS requirements with a distributed MAC protocol on a per flow basis for a reasonably loaded mesh.

3. Using the existing MAC in mesh

Prioritized access increases the probability of higher 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 is not always a simple multiple of the delay experienced for a single hop in a non-mesh environment. A single hop flow in a mesh may experience a longer delay than non-mesh experience would suggest. The prevalence of hidden nodes and the interaction of contention-based access with multi-hop flows can impose latency increases on both single and multi-hop flows beyond what non-mesh experience suggests.

3.1 Hidden node collisions

Collisions in wireless mesh networks occur for two reasons. One type of collision is caused by simultaneous transmissions by two or more devices located sufficiently close that their signals result in signal to interference plus noise ratio (SINR) at the receiver that is too low for proper decoding. Typically such a collision occurs if the backoff delay of two or more such devices waiting to transmit expires simultaneously. Obviously, the higher the concentration of active devices in the vicinity of a transmitter-receiver pair, the higher the collision rate observed.

Another way collisions arise is from 'hidden nodes' [27]. 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 of a hidden node is illustrated in Figure 1, where node F, which is outside the sensing range of node A, is a hidden node when node A transmits to node B. The sensing range of a transmitter refers to a range within which any node can sense the received signal, whose power level exceeds a sensing threshold. Collisions can result from hidden nodes as follows. If node B is within the interference range of node F and nodes A and F engage in overlapping transmissions, B will be unable to decode a transmission from A. This is known as a 'hidden node collision'.

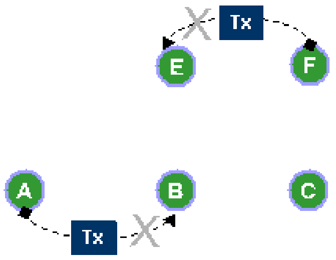

Figure 1. Hidden node collision

The IEEE 802.11 MAC protocol offers RTS/CTS and TXOPs as possible remedies for hidden node collisions [5]. RTS/CTS involves the use of a multiple-frame handshake between the transmitter and receiver, which comprises short control frames – namely, RTS (Request to Send) and CTS (Clear to Send) frames [25], [28]. The RTS, which is sent by the source of the pending transmission to head off a collision, 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 notifies also 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 sender of a frame cannot sense an ongoing transmission that its intended recipient hears, the CTS will not be sent; using the RTS will preempt a hidden node collision involving the frame. Once the frame transmission starts, any hidden nodes would refrain from transmission because they received the CTS, thus averting hidden node collisions.

A TXOP (Transmission Opportunity) also provides protection against hidden nodes. The frame initiating the TXOP carries a field indicating the TXOP duration, which is the period of time for which the channel is reserved. The receiver returns this information in the acknowledgement frame, which notifies the neighboring nodes of the channel reservation as it carries a field with the duration of the channel reservation.

It must be noted that neither RTS/CTS nor TXOPs reserve the channel for the transmission on the next hop of a multi-hop transmission. In the discussion that follows, we explain how this can be done through 'express forwarding', and how this capability can be used for the transmission of the RTS and for a TXOP, thus combining their respective benefits.

There is a tradeoff in using RTS/CTS. The penalties include the increased bandwidth taken by the control frames. Additionally, collisions are not entirely avoided. Both the RTS and CTS may be involved in collisions. The RTS may be involved in a regular collision or a hidden node collision, just like any other frame. The CTS may cause a hidden node collision to an ongoing transmission its sender cannot hear. This notwithstanding, the use of RTS/CTS pairs is advantageous if they avert collisions involving longer frames. A tradeoff exists, therefore, between the increased bandwidth taken by the control frames and the decrease in channel time lost to collisions.

Hidden node collisions are more prevalent in mesh networks than in WLANs. Hidden node collisions arise in WLANs, but not with the same frequency as in the mesh. In infrastructure WLANs -- that is, WLANs where stations communicate typically through the AP -- hidden node collisions occur only on uplink transmissions, as all devices can hear the AP. WLANs with overlapping coverage areas can avoid cross collisions by selecting different channels. Thus, co-channel WLANs could be separated by longer distances than possible for nodes of a single-channel mesh, avoiding cross collisions. Figure 2 illustrates hidden node collisions in WLANs. Simultaneous transmissions by nodes A and E to their serving AP at node B will fail, because A and E, although part of the same WLAN, cannot hear one another. On the other hand, simultaneous transmissions by nodes A and F (or by E and F) to their serving APs, at nodes B and D, respectively, will be received successfully if the two APs use different channels. If the nodes in Figure 2 represented a mesh, all operating on the same channel, simultaneous transmissions by nodes A and

F (or by E and F) to nodes B and D, respectively, would fail because of hidden node collisions.

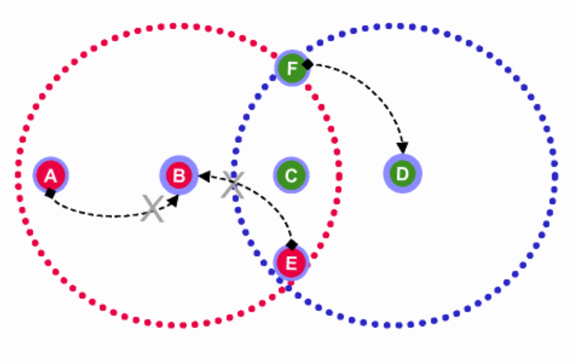

Figure 2. Hidden node collisions in WLANs

Hidden nodes are most prevalent in mesh networks used for range extension because the proportion of nodes that can hear each other is small. A node cannot typically decode the transmissions by neighbors of neighbor nodes. While long separation between communicating nodes gives rise to channel re-use potential across a mesh, the derived benefit disappears in a mesh using a single channel. Between a pair of potentially non-interfering nodes in a connected mesh, lies a third node that can cause interference to both pair members, operating on the same channel. This gives rise to hidden nodes and the potential of hidden node collisions. A grouping where all nodes can hear one another, i.e. a grouping without hidden nodes, will perform better because of the effectiveness of collision avoidance. Such a grouping, however, might be covered as well by a single WLAN.

Hidden nodes arise also when a single-channel mesh is located near a WLAN that uses the same channel. The mesh neighbors of a mesh node within sensing range of a WLAN device may be outside the sensing range of the same WLAN device and therefore become hidden nodes. Similarly, other WLAN devices would be hidden nodes for the mesh node closest to the WLAN.

The prevalence of hidden nodes increases collision rates and retransmissions, leading to higher channel utilization per attempted transmission and to dropped frames. Figure 1 illustrates common topologies in wireless meshes that cause repeated collisions and dropped frames. 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. These collisions are likely to repeat on re-transmission because nodes A and F cannot hear each other. The backoff delay of each is decremented in time regardless of whether the other is transmitting, and transmission is likely to be attempted while the other is transmitting, simply because they cannot hear each other. 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

Multi-hop flows in a single-channel mesh may experience repeated hidden node collisions along several of their hops, causing the end-to-end delay to build up. In addition, their interaction with contention-based access can cause latency increases on other single and multi-hop flows beyond what non-mesh experience suggests. This novel behavior of meshes can impact nearby WLANs as well.

Longer delays can be caused by multi-hop flows because of special features of the contention-based access mechanism. When a transmission is involved in a collision, it is at a disadvantage relative to transmissions attempted for the first time. 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. A retransmit backoff delay would typically be longer than the backoff delay drawn by a forwarding device, immediately following the successful receipt of a frame of a multi-hop flow. Therefore, if a re-transmitting device is in the vicinity of a multi-hop flow, this device may have to wait for the completion of multiple hops of that flow before retransmission is possible because of its longer retry backoff.

Collisions are often caused by a multi-hop flow as it advances along its path. Transmissions near a multi-hop path are vulnerable. The acknowledgement of successful receipt of a frame on one hop may collide with a transmission further down near the path, which will likely have to wait for the entire multi-hop flow to complete. Figure 3 illustrates how a multi-hop flow may delay a transmission near its path. The acknowledgement from node C to node D causes a collision for the transmission to node B. Node A will probably have to wait for nodes D and E to forward the frame they receive before it can retransmit because of its longer backoff.

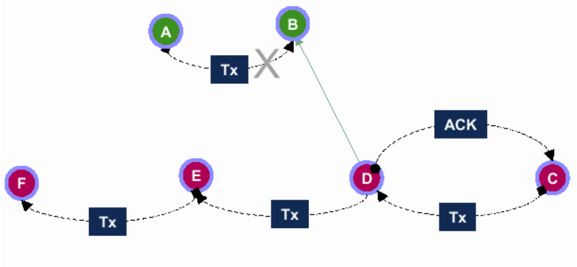

Figure 3. Collision due to multi-hop flow

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

Three measures are proposed 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, according to the first measure, the likelihood of averting a repeat collision increases for two nodes whose transmissions collided because they cannot hear each other. In this case, the backoff delay represents the clock time -- not the cumulative channel idle time – each such node will wait before transmitting, as the transmission of the other node is not heard. Therefore, increasing the retry contention window increases the probability that the transmissions of the two nodes will not overlap in time.

This measure can be implemented simply when using the IEEE 802.11 MAC protocol by allowing the contention window size for backoff delay to increase more. The default values for CWmax, the contention window size in the IEEE 802.11 Standard such that once it is reached, the window size is no longer doubled after a collision can be raised.

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. Because it uses carrier sense functions and the collision avoidance backoff mechanism, it can interoperate with WLANs using the same channel. A high-level overview of express forwarding was first given in a presentation to the IEEE 802.11s task group [29].

According to express forwarding, multi-hop transmissions are expedited by reserving the channel via the transmitted frame on each leg of the multi-hop path for the next hop. The notion of an Express Forwarding TXOP (EF-TXOP) thus arises, which is a time-space extension of the IEEE 802.11 TXOP. Transmit opportunities (TXOPs) enable a source to transmit multiple frames following a single successful channel access attempt, without having to contend for the channel. That is, a source transmits consecutive frames from the same access category without the need to contend (i.e. engage in backoff) more than once. In an EF-TXOP, consecutive linked transmissions of a multi-hop flow are made without the need to contend more than once. Reservation is done the same way as for TXOPs. 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 on a multi-hop path.

Reservation of the channel for an EF-TXOP is done through the virtual carrier sense mechanism used in IEEE 802.11 devices, as in the case of the TXOP, Virtual carrier sense is one of two mechanisms that enable a device to keep track of the activity level of the channel. Physical carrier sense is based on the receiver detecting energy in the channel. Virtual carrier-sense relies on a timer, referred to as the network allocation vector (NAV), which indicates how long the medium will be busy. A node is not allowed to transmit while its NAV timer is set. The NAV is set and updated based on the Duration field value contained in transmitted frames. The response frame, which is 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. Thus the duration field and the NAV timer provide a means for channel reservation.

The channel is reserved for a TXOP by setting the duration value of a frame long enough to cover at least one additional frame and its response frame, and by waiting a shorter time between transmissions than any other source contending for the channel. The Duration field of the response frame thus indicates the length of the following frame in the TXOP or the remaining TXOP duration. Because all but one frame in a TXOP is transmitted without contention, TXOPs help reduce the frequency of collisions. This increases channel use efficiency.

When a frame is express forwarded, the channel is reserved by extending the Duration field value of the frame long enough to silence all neighboring nodes and give the receiving node the opportunity to seize the channel and forward the frame. As illustrated in example of a three-hop flow in Figure 4, the NAV timer at neighboring nodes is set according to the Duration field value on a frame that is to be express-forwarded on the next hop. The Duration field value is longer than the time period the channel is occupied by the transmission and acknowledgment of the frame for the first two of the three hops of the path illustrated. This way, following the contention for the transmission on the first hop, an express-forwarded frame is transmitted quickly on the second and the third hop without contention, causing the multi-hop end-to-end delay to decrease. As in the case of a TXOP, EF-TXOPs help reduce collisions and thus increase 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 jitter for non-express-forwarded traffic.

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 their NAV is set 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.

Express forwarding can be used for the transmission of the RTS and for a TXOP, thus combining their respective benefits.

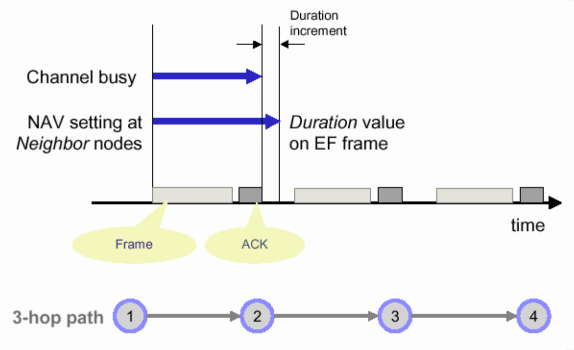

Figure 4. Express Forwarding reservation

4.2.1. Combining TXOPs and EF-TXOPs

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. This can be achieved through the duration field of any of the frames in the TXOP. It suffices to extend the duration field of the last frame transmitted in the TXOP. If multiple express-forwarded frames are part of a TXOP, they can all be express-forwarded by the receiving node only if they are all going on the same link next. If the express-forwarded frames of a received TXOP request different next-hop destinations, the receiving node will have to select one mesh neighbor for its upcoming EF-TXOP. It may have to initiate other EF-TXOP(s) for the remaining packets that requested express forwarding.

When the received express-forwarded frame must be forwarded, if other frames queued at the receiving node can be sent in the same TXOP (that is, meets existing TXOP restrictions), 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.

Embedding TXOPs within EF-TXOPs increases the efficiency of channel utilization. Express forwarding requires all nodes in the vicinity of the source to wait for the received frame to be processed at the IP layer and returned to the MAC layer for forwarding. This may cause the channel to remain unused if the time required for processing is longer than the time for transmitting the acknowledgment, assuming processing and acknowledgment is done in parallel. Transmitting more frames in the same TXOP following an express-forwarded frame allows the channel to be used while the express-forwarded frame is processed at the receiving node. Transmitting frames that are queued at the receiving node ahead of a received express-forwarded frame, in the same TXOP, allows the channel to be used while the received express-forwarded frame is processed for forwarding. Using either approach to place express-forwarded frames in a TXOP can prevent the channel from sitting unused.

4.2.2. RTS/CTS with Express Forwarding

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. The two mechanisms complement each other and can be used together.

Express forwarding can be used to send an RTS along each of the legs of a multi-hop path. The duration field of the RTS will reserve the channel not only for the protected transmission by the source of the RTS, but also for the next RTS transmitted by the forwarding node. 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

The retransmission of an express-forwarded frame that has been involved in a collision can also be expedited. Retransmission of a failed transmission typically involves contending with a backoff delay drawn 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 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. It is less likely for the two retransmissions to overlap in time since 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 then 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.

4 Performance evaluation

The performance benefits of express forwarding and express retransmission have been demonstrated in several studies for a range of scenarios [30], [31].

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 various channel access scenarios. We present here results from one of the studies, which 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. Single-hop flows use EDCA for all scenarios. In the second scenario, express forwarding is employed for the multi-hop flows. In the third scenario, the multi-hop flows use express forwarding and express retransmission.

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 multi-hop flows, with three hops each, 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 5. The IP phones generate bi-directional streams communicating either with mesh peers or with the outside world through Node 0, which is the mesh portal. There was no node mobility; hence, static routing is employed. Table 2 presents the key traffic and MAC parameters.

Table 1. Scenario description

Scenario	Description
1. EF Disabled	Express Forwarding disabled
2. EF Enabled	Express Forwarding enabled for multi-hop flows
3. EF-ERTX Enabled	Express Forwarding & Express Retransmission enabled for multi-hop flows

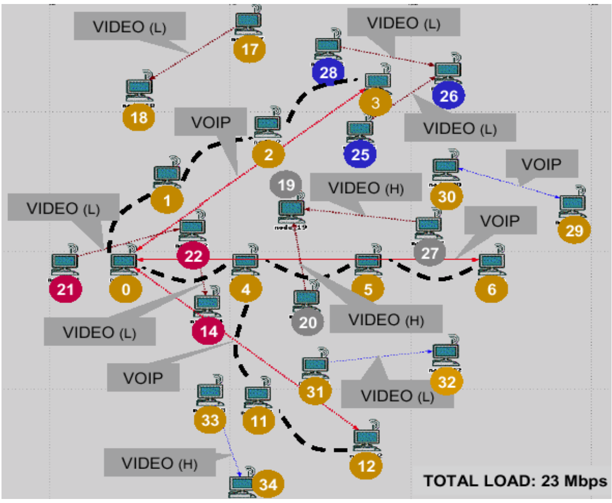

Figure 5. Network layout

Table 2. Key traffic and MAC parameters

Traffic Type	Payload (bytes)	Frame Spacing (ms)	CWmin*	CWmax** WLAN/ Mesh
VoIP call	200	20	7	15/ 1023
Low-resolution Video	1464	8	15	31/ 1023
High-resolution Video	1464	2.83	15	31/ 1023

*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

All nodes were equipped with a single 802.11a radio. The channel was assumed to be noise free. Application data traffic was transmitted at 54 Mbps and acknowledgments at 24 Mbps. A 50 μsec IP processing delay was assumed at each node, typical delay for for processors in devices now being implemented. Processing of a frame starts as soon as it is received and in parallel with the transmission of an acknowledgement.

Simulations were conducted by using the OPNET Modeler modeling platform [32]. Statistics were computed over a simulation time of two minutes, starting when steady state was reached. Repeated experiments, obtained by varying the starting time of the flows randomly, showed negligible change in the measured statistics.

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 6 and 7 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.

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.

Table 3. Mean end-to-end delay (msec)

Scenario		EF Disabled	EF Enabled	EF-ERTX Enabled
Flow	Network			
Node 0 – Node 3 (M)	Mesh	22	5	2
Node 3 – Node 0 (M)	Mesh	19	3	2
Node 0 – Node 6 (M)	Mesh	2,698	8	3
Node 6 – Node 0 (M)	Mesh	2,562	4	3

Node 0 – Node 12 (M)	Mesh	3,583	17	6
Node 12 – Node 0 (M)	Mesh	3,448	16	7
Node 17 – node18 (S)	Mesh	12	4	3
Node 29 – Node 30 (S)	Mesh	9	3	3
Node 30 – Node 29 (S)	Mesh	4	3	2
Node 31 – Node 32 (S)	Mesh	8	4	3
Node 33 – Node 34 (S)	Mesh	28	14	7
Node 20 – Node 19 (S)	WLAN 1	6	4	4
Node 27 – Node 19 (S)	WLAN 1	8	5	5
Node 21 – Node 22 (S)	WLAN 2	4	3	3
Node 22 – Node 14 (S)	WLAN 2	3	2	2
Node 25 – Node 26 (S)	WLAN 3	3	2	2
Node 28 – Node 26 (S)	WLAN 3	3	2	2

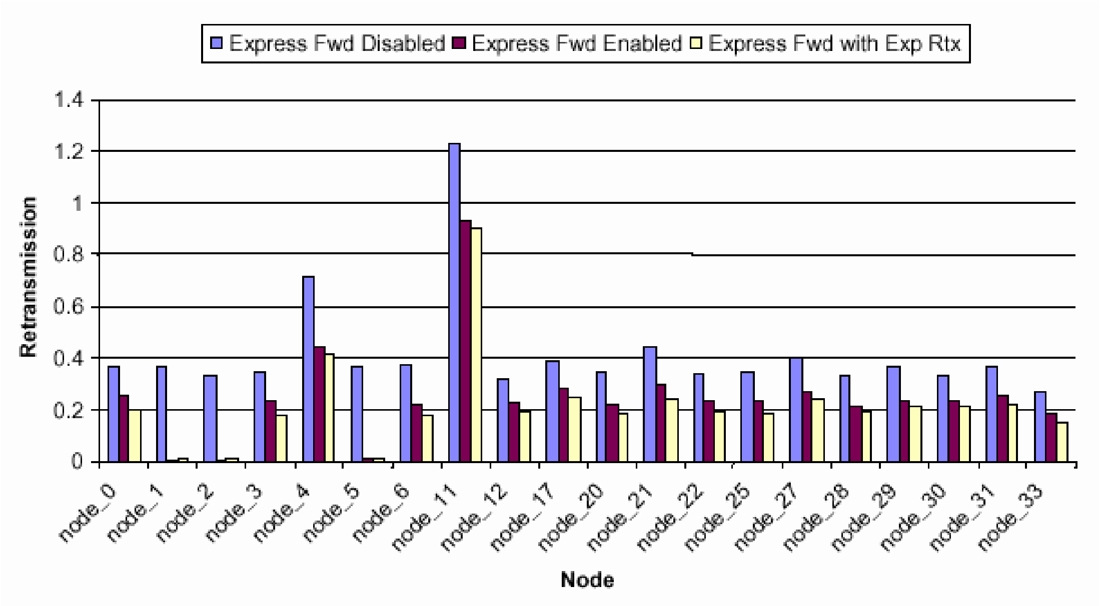

Figure 6. Normalized retransmissions by node

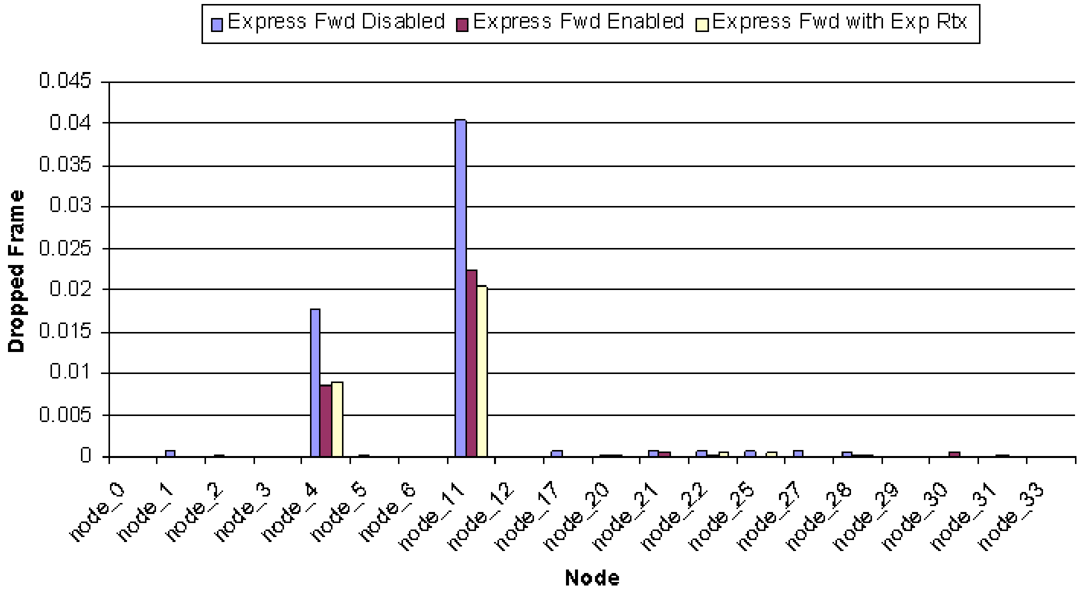

Figure 7. Normalized dropped frames by node

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.

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. The standard IEEE 802.11 EDCA access mechanism could not meet these requirements.

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. Regardless of the criterion used to establish fairness, however, 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. The benefit derived from it will depend on the MAC protocol used for channel assignment and scheduling radio use. This would be the subject of future investigation.

References

[1] M. Benveniste, "A Distributed QoS MAC Protocol for Wireless Mesh", Sensor Technologies and Applications, 2008. SENSORCOMM '08. Second International Conference, Aug. 2008, pp. 788 – 795.

[2] C. Perkins, editor. *Ad Hoc Networking*. Addison-Wesley, 2001.

[3] I.A.Akyildiz and X. Wang, "A survey of wireless mesh networks", *IEEE Radio Communications*, September 2005.

[4] H. Zhai, J. Wang, X. Chen and Y. Fang, "Medium access control in mobile ad hoc networks: challenges and solutions", Wireless Communications and Mobile Computing, Vol 6, 2006, pp. 151 –170.

[5] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical (PHY) Layer Specifications, ANSI/IEEE Std 802.11, 2007 Edition

[6] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing", in *Proceedings of the* 2nd *IEEE Workshop on Mobile Computing Systems and Applications*, (New Orleans, LA), February 1999, pp. 90-100.

[7] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers", in *ACM SIGCOMM' 94 Conference on Communications Architectures, Protocols and Applications*, 1994, pp. 234.244.

[8] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks", in *Mobile Computing* (Imielinski and Korth, eds.), vol. 353, Kluwer Academic Publishers, 1996.

[9] S. Toumpis and A. J. Goldsmith, "Performance, Optimization, and Cross-Layer Design of Media Access Protocols for Wireless Ad Hoc Networks," in Proc. IEEE International Conference on Communications, May 2003, pp. 2234-2240.

[10] J. Chen and S.Hsia, "Joint Cross-Layer Design for Wireless QoS Video Delivery," in Proc. International Conference on Multimedia and Expo, July 2003, pp. 197-200.

[11] C. L. Gwee, Y. Qin and W. K. G Seah, "Bundled Virtual Circuit: A Proposed Cross Layer Design of Routing and Scheduling for QoS Services in MANETs" VTC Spring 2006, pp. 1283-1287.

[12] *Recommendation ITU-T G.114, One-Way Transmission Time,* Int'l Telecommunication Union, Geneva, 1996.

[13] M. Benveniste, "WLAN QoS", in *Emerging Technologies in Wireless LANs* (B. Bing, ed.), Cambridge Univ. Press, 2008.

[14] A. Nasipuri, and S. Das, "A Multichannel CSMA MAC Protocol for Mobile Multihop Networks." In Proc. of IEEE WCNC '99.

[15] S-L Wu, C-Y Lin, Y-C Tseng, and J-P Sheu, "A new multi-channel MAC protocol with on-demand channel assignment for mobile ad hoc networks", International Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN) 2000, pp. 232– 237.

[16] Y-C Tseng, S-L Wu, C-Y Lin, and J-P Sheu. A multi-channel mac protocol with power control for multi-hop mobile ad hoc networks. In *Proceedings of 21st*

International Conference on Distributed Computing Systems Workshops, April 2001, pp. 419–424.

[17] A. Tzamaloukas, and J. J. Garcia-Luna-Aceves; .A Receiver-Initiated Collision-Avoidance Protocol for Multi-channel Networks.; Infocom '01.

[18] L. Bao, J. and J. Garcia-Luna-Aceves, “Hybrid channel access scheduling in ad hoc networks”, in Proc. IEEE ICNP’02, November 2002.

[19] N. Choi, Y. Seok , and Y. Choi, “Multi-channel MAC protocol for mobile ad hoc networks”, in Proc. IEEE VTC 2003-Fall, October 2003.

[20] K. Liu, T. Wong, J. Li, L.Bu, and J. Han, “A reservation-based multiple access protocol with collision avoidance for wireless multihop ad hoc networks” in Proc. IEEE ICC’03, May 2003.

[21] T. You, C. Yeh, and H. Hassanein, “A new class of collision prevention MAC protocols for wireless ad hoc networks”, in Proc. IEEE ICC’03, May 2003.

[22] Y. Liu, and E. Knightly, “Opportunistic Fair Scheduling over Multiple Wireless Channels”, in Proc. IEEE INFOCOM '03.

[23] J. So, and N. Vaidya, “Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using A Single Transceiver”, in Proc. MobiHOC 2004.

[24] M. Benveniste, “Tiered Contention Multiple Access (TCMA), a QoS-Based Distributed MAC Protocol”, in Proc. PIMRC, Lisboa, Portugal, September 2002.

[25] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical (PHY) Layer Specifications, ANSI/IEEE Std 802.11, 1999 Edition.

[26] L. Kleinrock and F. A. Tobagi. Packet switching in radio channels: Part–I - carrier sense multiple access modes and their throughput-delay characteristics. *IEEE Transactions in Communications*, COM-23(12), 1975, pp.1400–1416.

[27] F. A. Tobagi and L. Kleinrock. Packet switching in radio channels: Part–II - the hidden terminal problem in carrier sense multiple-access and the busy-tone solution. *IEEE Transactions in Communications*, COM- 23(12), 1975, pp.1417–1433.

[28] P. Karn. MACA: A new channel access method for packet radio. In *Proceedings of ARRL/CRRL Amateur Radio* 9th *Computer Networking Conference*, 1990.

[29] M. Benveniste, “‘Express’ Forwarding for Single-Channel Wireless Mesh”, IEEE Doc 802.11-07-2452r2.

[30] M. Benveniste and K. Sinkar, “Performance Evaluation of ‘Express Forwarding’ for a Single-Channel Mesh”, IEEE Doc 802.11-07-2454r1.

[31] M. Benveniste and K. Sinkar, “More on Performance Evaluation of 'Express Forwarding' for Mesh”, IEEE Doc 802.11-08-0142r0.

[32] OPNET Modeler, http://opnet.com/solutions, May 2008.