Block Negative Acknowledgement protocol for reliable multicast in IEEE 802.11

Yousri Daldoul a, b, Djamal-Eddine Meddour a, b, Toufik Ahmed b, Raouf Boutaba c

a Orange Labs Network, Lannion, France
b Bordeaux I University, Bordeaux, France
c David R. Cheriton School of Computer Science, University of Waterloo, Canada

Abstract

The default multicast transport of the IEEE 802.11 standard does not use any feedback policy to detect and retransmit missing packets. Consequently, it has a limited reliability. In this paper we introduce the Block Negative Acknowledgment (BNAK) protocol as a solution for a reliable multicast transport in wireless networks. Using BNAK, the Access Point (AP) transmits a block of multicast packets followed by a Block Negative Acknowledgement Request (BNR). Upon the reception of a BNR, a multicast member sends a BNAK response, only if it has some missing packets. A BNAK is acknowledged and therefore is delivered reliably to the AP. Moreover it is transmitted after channel contention in order to avoid eventual collisions with other feedbacks. Under the assumption that (1) the receiver is located within the coverage area of the sender, (2) the multicast packets are delivered using the appropriate data rate and (3) the collisions are avoided, the Packet Error Rate (PER) of the network becomes very low. To guarantee a limited PER, BNAK requires the use of a collision prevention feature (such as CTS-to-Self), and defines a retirement/re-activation procedure. Thus few feedbacks are generated and the bandwidth is saved. We show that our protocol has very high scalability and outperforms the proposals defined by 802.11v and 802.11aa considerably. Particularly our protocol can achieve a throughput exceeding 10 times that of GCR-BACK.

Article info

Article history:
Received 26 May 2015
Received in revised form 16 June 2016
Accepted 12 July 2016
Available online 25 July 2016

Keywords:
Block Negative Acknowledgment
Reliable Multicast
WLAN
802.11aa
802.11v

1. Introduction

The main advantage of multicast transport is its high scalability (Miroll et al., 2010): a multicast packet is sent once single time regardless of the number of the multicast receivers. However, the conventional multicast protocol of 802.11 (Wireless LAN, 2012a) is not reliable due to the absence of any feedback policy. Thus, missing packets are definitely lost. In particular, multicast transmissions are vulnerable to collisions (Sun et al., 2002) which are a principal loss factor in WLANs, and to device unavailability. Besides, multicast packets are by default delivered at the lowest data rate to reach the entire coverage area of the wireless network. This selected rate wastes the bandwidth and reduces the network throughput. The selection of a higher transmission rate is not obvious since the Access Point (AP) is not aware about the reception capabilities of the group members. To resolve both of the unreliability issue and the reduced efficiency problem, it is necessary to define a feedback policy for the multicast transport.

The IEEE 802.11v (Wireless LAN, 2011) and 802.11aa (Wireless LAN, 2012b) are two recent amendments which enhance the reliability of multicast transmissions. The former defines Directed Multicast Service (DMS) while the latter introduces the Groupcast with Retries (GCR) service. DMS converts a multicast stream into multiple unicast sessions. It resolves the unreliability issue on the expense of bandwidth. Therefore, DMS has a very limited scalability. On the other hand, GCR defines the Block Ack (BACK) policy which allows the AP to recover the feedback of each member using the individual Block Ack Request (BAR)/BACK exchange. Similar to DMS, GCR-BACK is a reliable multicast protocol. But it incurs an overhead which depends on the group size. Thus, the scalability of GCR-BACK is also limited. Besides, GCR defines Unsolicited Retry (GCR-UR). This policy does not use feedbacks, but it transmits each packet several times in order to improve the reliability. Therefore, the efficiency of GCR-UR decreases significantly as the retry count increases. Pseudo-broadcast protocols and collision avoidance approaches are also proposed to overcome reliability issue for multicast flows, yet, obtained results were not completely satisfactory. On the other hand, there were several efforts that intend to tackle multicast transport over 802.11 networks (Wireless LAN, 2011, 2012b; Sun et al., 2002; Campolo et al., 2009; Tanigawa et al.,...
In this paper we introduce a new multicast protocol called Block Negative Acknowledgement (BNAK) in order to alleviate the aforementioned limitations in the related standards and the state of the art. A key requirement underlying the design of BNAK is to define a reliable and a scalable multicast transport in 802.11 networks while retaining the compatibility with the former and the newer amendments of the 802.11 standard. The main design principle of our protocol relies on the fact that losses in WLANs become very limited if the appropriate actions are taken. As such, BNAK requires the use of a collision prevention feature (such as CTS-to-Self), and defines a retirement/reactivation procedure to guarantee a limited PER. Then, it recovers negative acknowledgements from members experiencing packet losses. As the principal loss factors are avoided, a very limited number of feedbacks are transmitted and the bandwidth is saved. These feedbacks are also useful to select the most appropriate transmission rate. BNAK operations include new mechanisms for protection against useless transmission, member retirement and reactivation, transmission procedure and group management. We evaluate the performance of BNAK using both analytical and simulation approaches with a variety of parameters. The obtained results demonstrate that BNAK surpasses largely the existing solutions with respect to the reliability and throughput.

To summarize, the main contributions of this paper are as follows. First, we define a membership management function at the MAC layer. Second, we design a new multicast protocol, called BNAK, for 802.11 networks. Third, we define a retirement/reactivation procedure for members experiencing temporal channel fluctuation. Fourth, we define an analytical model to evaluate the throughput and the scalability of BNAK and GCR-BACK. Fifth, we validate the defined model and we evaluate our protocol using extensive simulations.

The remainder of this paper is organized as follows. Section 2 provides an overview of related works. In Section 3 we present the different components and operations of our proposal BNAK. We devote Section 4 to present our analytical model. We evaluate BNAK in Section 5. Finally we conclude in Section 6.

2. Related works

Many protocols have been proposed for improving the reliability and performance of multicast transport over 802.11 networks (Wireless LAN, 2011, 2012b; Sun et al., 2002; Campolo et al., 2009; Tanigawa et al., 2010; Kuri and Kasera, 1999; Choi et al., 2007, 2010; Lim et al., 2012; Miroll et al., 2010; Chandra et al., 2009; Santos et al., 2010; Daldoul et al., 2013, 2012; Shin et al., 2011). Yet, all of them suffer of one or several of the following limitations: low reliability, limited scalability, reduced efficiency and no compliancy with 802.11.

Batch Mode Multicast MAC (BMMM) (Sun et al., 2002) requires the acknowledgement of all the multicast members. The transmission procedure of this protocol is as follows. Initially, the AP exchanges RTS/CTS with all the members, then sends the multicast packet. Finally it sends a Request for ACK (RAK) to each receiver. If a member receives the data packet correctly, it replies to the RAK with an ACK. Otherwise, it does not reply.

The major weakness of BMMM is its limited efficiency (i.e. throughput). Indeed, BMMM requires the exchange of several control packets with all the group members. The number of these packets depends on the group size which limits the protocol scalability. Besides, the multicast delivery to large groups requires an important transmission slot. This impacts the quality of time sensitive applications sharing the channel. Another limitation of BMMM is that this protocol is defined for individual acknowledgements and does not support block transmissions.

Reliable Access Multicast Protocol (RAMP) (Campolo et al., 2009) slightly enhances the performance of BMMM by considering the following features: modified RTS called MRTS, and the encapsulation of the multicast flow into a new packet called Multicast DATA (MDATA). Therefore, one single MRTS is enough to receive Multicast CTS (MCTS) sequentially from all the group members. Similarly, MDATA is sufficient to recover all the feedbacks. Moreover, MRTS/multiple-MCTS handshaking is dynamically enabled and disabled according to loss history. The overhead of RAMP is a function of the group size. Therefore, this protocol has a limited scalability. Besides, it does not support the block transmissions. Furthermore, the sequential reply is the main novelty of RAMP to reduce the overhead. However, this feature is not appropriate for wireless networks. This is because 802.11 requires that a station should defer its transmission for a long period called extended inter-frame space (EIFS) following any reception failure. This delay prevents the members to acknowledge at the appropriate instant if they fail to receive the last feedback correctly. In this case, the multicast packet is retransmitted even if it has been received correctly. We note that the multicast receivers are not always able to hear their respective transmissions since they may be located at the opposite sides from the sender. Therefore many useless retransmissions may occur. This limits the efficiency of RAMP significantly.

The authors of (Tanigawa et al., 2010) propose a transparent multicast/unicast translation method which converts multicast packets into multiple unicast flows at the MAC layer. Therefore, the multicast traffic is delivered reliably at the expense of bandwidth, limiting the scalability of the proposed method.

2.2. NAK-based protocols

Negative Acknowledgement (NAK) protocols are mainly designed to improve the scalability of Multicast applications. Many of these protocols belong to the Transport layer, while others operate at the MAC layer. For some access networks, such as Ethernet, there is no need to define a reliable MAC protocol: for these networks, the loss rate is very low (no losses related to collisions and path-loss), and protocols of the transport layer are enough to provide a reliable communication. But 802.11 networks operate over a lossy channel (collisions and path-loss are significant loss factors) and require a reliable MAC protocol in order to optimize the performance of the transport layer. For example, TCP will experience significant performance issues over WLAN without a reliable MAC protocol. For this reason, 802.11 defines a reliable unicast scheme. Similarly, multicast NAK protocols that belong to the transport layer will experience significant performance issues over WLAN without a reliable MAC-protocol.
2.2.1. Transport-layer NAK protocols

Among the most known Transport-layer NAK protocols, we find NORM (NACK-Oriented Reliable Multicast (NORM), 2009) and PGM (PGM Reliable Transport Protocol Specification, 2001), which operate like follows. The sender transmits packets and FEC repairs. Using the packet sequence numbers and a timer, the receiver is able to send a NAK to request any missing data. These protocols do not address any specific issues related to the lossy nature of the wireless networks. Without the help of a reliable MAC-layer protocol, they will experience very limited performance over IEEE 802.11 networks.

2.2.2. MAC-layer NAK protocols

Many NAK protocols are defined at the MAC layer of 802.11. In fact, these protocols select one single group leader that acknowledges successfully received packets. The other receivers are only allowed to send NAKs upon reception failures. The NAK intends to collide with the leader’s ACK and to destroy it: this is called “Feedback Cancellation” (Miroll et al., 2010). Once the ACK is lost, the sender retransmits the unacknowledged packet. However, the Feedback Cancellation is not compliant with the collision avoidance policy of IEEE 802.11.

This feedback principle is investigated by many works. One of them is Leader Based Protocol (LBP) [Kuri and Kasera, 1999]. The main characteristic of LBP is that it dedicates one single acknowledgement slot time for all the members, regardless of the multicast group size. In this way, the protocol ensures the scalability of the multicast service. Besides, LBP requires the use of RTS/CTS before each transmission. These control packets are exchanged with the group leader and limit the collisions. They are also required to advertise the multicast transmission and its duration. Therefore, if the multicast packet is lost, the group members are able to build a NAK and to send it at the appropriate instant.

However, the use of RTS/CTS increases the transmission overhead and reduces the efficiency of LBP. To alleviate this issue, Leader-based Multicast with Auto Rate Fallback protocol (LM-ARF) (Choi et al., 2007) proposes the use of CTS-to-Self protection instead. This protocol outperforms LBP since the overhead is reduced. However, both protocols still have a reduced efficiency since they do not support the block transfer nor the packet aggregation feature.

In (Lim et al., 2012), authors try to adapt the NAK principle to the aggregation transmissions. Using this approach, when the leader’s Block ACK (BACK) is missing, the multicast source requests an individual BACK from each member. We note that even if the packet loss rate is limited, increasing the number of aggregated packets and the multicast group size leads to an increased probability to have at least one NAK transmission. Therefore it is likely that the proposed scheme requests frequently individual BACKs following the already wasted time to send the NAK. This protocol inherits the reduced efficiency of the individual feedbacks.

This concept is very weak and its major issues are:

- The leader’s ACK may hide a NAK (Miroll et al., 2010) (in this case the missing packet is not retransmitted)
- No reliable method to detect packet losses: NAKs are built using the data received within erroneous packets which have bad CRC
- The Feedback Cancellation is not compliant with the collision avoidance policy of 802.11.

2.2.3. Novelties of BNAK

In this paper we define a new multicast protocol called BNAK that belongs to the NAK family and operates at the MAC layer. It uses the NAK principle in a completely new way to resolve issues experienced by former proposals. Particularly, our protocol does not use the Feedback Cancellation and is therefore compliant with 802.11. Besides, it provides a reliable method to detect packet losses, and is suitable for the recent packet aggregation futures. As BNAK belongs to the MAC layer, it can be combined with a Transport-layer NAK protocol (such as PGM (Daldoul et al., 2015)) in order to provide a reliable and scalable multicasting solution over large delivery networks.

2.3. Pseudo-broadcast protocols

Pseudo-broadcast protocols select one member to send a feedback following the successful reception of each multicast packet. The other members, however, do not send any notification. Thus, these protocols are scalable and compliant with the standard but they are not fully reliable. Besides, they do not allow the AP to select the appropriate data rate since the sender is aware about only one downlink. This leads the AP to select the lowest transmission rate which reduces the overall throughput.

In (Chandra et al., 2009), authors proposed the DirCast system to enhance the QoS of multicast services over IEEE 802.11 networks. DirCast selects one member called “target client” for each multicast group, then sends multicast packets using unicast to the selected client. The other clients receive the packets by monitoring the channel in the promiscuous mode. This solution does not consider the wireless medium characteristics and the different data rate capabilities of the multicast members; if the packets are transmitted at 54 Mbps, all receivers which do not support this rate or are located out of the coverage of 54 Mbps (but within the coverage of lower data rates) will miss the packets.

Leader-Based Multicast Service (LBMS) (Choi et al., 2010) does neither convert a multicast stream into unicast nor configure the receivers into the promiscuous mode. However, the selected leader is responsible to acknowledge each successfully received multicast packet. This protocol does not require an important overhead. Therefore it is scalable. However, as the authors acknowledge, LBMS is not fully reliable but is a semi-reliable protocol. Thus, it does not guarantee the quality of the multicast service.

2.4. Collision prevention

In 802.11 networks, collisions constitute a principal factor of packet losses. Therefore, protecting the multicast packets from collisions limits the loss rate and improves the reliability of the multicast transmissions. In this perspective, the authors in (Santos et al., 2010) propose to transmit multicast packets one PIFS period after the acknowledgement of any unicast packet. This is an efficient way to avoid the collision of the multicast packet but has one major weakness: the number of multicast transmissions will depend on the number of transmitted unicast packets. Therefore, if the unicast packet rate is less than that of the multicast one, multicast packets will be buffered even if the bandwidth is available, and may be rejected because of queue overflow.

Another way to prevent collisions is to transmit multicast packets during Collision Free periods following the Beacons. We note that this policy is used if at least one associated station is in the Power Save (PS) mode. This is an efficient way to reduce the packet loss rate but increases the latency of time-sensitive unicast applications sharing the medium with the multicast stream (Shin et al., 2011). Moreover it increases the power consumption of devices in the PS mode (He et al., 2008) since these receivers are required to wait multicast transmissions before requesting their buffered packets.

In (Daldoul et al., 2013), we proposed the Busy Symbol (BS)
mechanism as an efficient solution for collision prevention. To protect multicast packets against collisions, the AP transmits a short OFDM symbol, called BS, before transmitting the multicast packet itself. The main idea of using the BS is to briefly occupy the channel, so that the contending stations, which are sensing the channel, defer their transmissions. Therefore, if no station has started transmission along with the BS, this symbol will create a contention free medium for a period of time equal to at least DIFS period after the BS end. This delay allows the AP to transmit the multicast packet without collision. If the medium is sensed to be busy after the transmission of BS, the AP defers the multicast transmission and avoids the collision.

2.5. IEEE 802.11v

The IEEE 802.11v amendment defines the Directed Multicast Service (DMS) in order to resolve the unreliability issue of multicast transport. This service allows a receiver to request the AP to send a multicast traffic as individually addressed packets (i.e. using unicast transport). This ACK policy guarantees the same unicast reliability degree to individually addressed multicast flows at the expense of bandwidth. Hence, DMS can be used to stream a standard traffic to a group of limited size, but does not scale well for High Definition (HD) streams like HDTV. In a previous work, we showed that losses caused by queue overflow may significantly exceed losses caused by the wireless medium (Daldoul et al., 2012). Moreover, the DMS does not allow the use of block transfer. We note that IEEE 802.11v defines the required procedure to establish DMS sessions but does not define any functionality to manage multicast groups and to identify their members.

2.6. IEEE 802.11aa

The IEEE 802.11aa standard defines the required MAC enhancements to provide a robust audio and video streaming. Among others, it proposes the Reliable Groupcast with Retries (GCR) service to improve the reliability of multicast transmissions. GCR provides two retry policies: GCR-Unsolicited-Retry (GCR-UR) and GCR-Block-Ack (GCR-BACK).

With GCR-Unsolicited-Retry policy, the multicast source defines a retry limit, say ‘N’, and transmits each multicast packet ‘N’ times without waiting for any feedback after each transmission. The retransmission of the same packet several times allows the sender to increase the probability of successful delivery. However, the use of this policy does not guarantee the QoS and generates a significant overhead.

GCR-BACK policy is similar to the basic block transfer of unicast. According to this policy the sender transmits a block of multicast packets followed by multiple exchanges of Block Ack Requests (BAR) and feedbacks. A member is allowed to reply only upon the reception of an explicit request. The received feedbacks allow the AP to detect missing packets. These packets are retransmitted until their lifetime limit is reached. Therefore, the GCR-BACK guarantees the same reliability degree of the unicast transport. Besides, the 802.11aa requires the use of a protective mechanism (like RTS/CTS, CTS-to-Self....) to avoid collisions.

CTS, data packets, BARs and BACKs within a block transfer are separated with a SIFS period. If the medium remains idle within a period of PIFS (i.e. SIFS plus one SlotTime) after the transmission end of a BAR, the AP concludes the reception failure of the last BAR and sends it again immediately. The AP retransmits a BAR in this way until it detects a transmission before the PIFS expiry or the lifetime of all the multicast packets expires. If the AP detects the BACK transmission but does not receive it correctly, then the AP retransmits the BAR following channel contention.

The GCR service does not define any procedure to manage the multicast groups. However, it proposes the use of IGMP snooping (Christensen et al., 2006) to achieve group membership detection and management.

3. Block Negative Acknowledgement (BNAK)

This section describes the main operations of BNAK. We first describe the group membership management function and then detail the transmission procedure in Sections 3.1 and 3.2 respectively. In Section 3.3, we show how to optimize BNAK utilization to avoid useless transmissions. Member reactivation and retirement procedures are outlined in Section 3.4.

3.1. Group management

In our design of BNAK, we consider that only the AP is required to gather information about group members. Therefore a multicast member does not need to perform any snooping at the MAC layer. This minimizes the client requirements and simplifies the deployment of our protocol. Basically, when a client joins a multicast service, the AP is responsible for notifying the member’s MAC layer. Similarly, when the client leaves the group, the AP sends a message to the client’s MAC layer in order to remove this station from the multicast group.

Most AP devices integrate both the AP and the router functionalities within the same terminal. Therefore membership information is available at the Network layer of the same equipment. In order to ensure the awareness of the MAC layer about group membership, we propose the exchange of internal messages (packets or signals), called membership notifications, between the Network and the MAC layers of the AP equipment. These messages are transmitted from the router level to the 802.11 MAC layer, whenever a new event occurs.

Two possible events may occur: (1) a new member joins the group and (2) a member leaves. As the notification messages are internal to the same device, they do not need any bandwidth resources and they are transmitted reliably to the MAC layer. Hence the AP does not need to acknowledge them. We highlight that this group management procedure eliminates the need for snooping IP packets at the MAC layer, and therefore reduces the device processing load compared to a snooping-based approach. The group membership management function may be achieved in a constructor dependent fashion when the AP equipment is entirely built by the same manufacturer. But in most cases, the chipset driver and the network stack are implemented by different parties. In this case it is necessary to standardize the notification interface between the network and the MAC layers. The signal-based notification method is the most appropriate one and is commonly used to configure the MAC layer and to set the different parameters such as: CWmin, CWmax, Retry limit, packet lifetime limit, beacon interval, etc. However, this method depends on the operating system (Windows, Linux, Android, embedded systems, etc.). On the other hand, the approach based on the internal packet exchange is another alternative which runs transparently to the underlying system architecture. In this paper we provide a simple, yet an accurate example of the notification function using the packet exchange method. This method may be easily converted later to the signal-based approach.

When the Network layer of the AP receives an IGMP/MLD join or leave request, it sends a notification packet to the MAC layer. Similar to the Address Resolution Protocol (Plummer, 1982), the notification packet uses a simple message format and is encapsulated in an Ethernet frame. To identify this packet, we use a specific EtherType of 0x0000. This value is not attributed and will not cause any processing confusion. However we neither add
paddi ng nor CRC to the packet since it does not leave the machine. Besides, we recommend the use of the MAC address of the AP as the source and the destination addresses of the Ethernet frame. This allows the MAC layers, which do not support BNAK, to delete the packet or to send it back to the Kernel. The format of the Ethernet notification packet is illustrated in Fig. 1.

The MAC address of the member allows the AP to determine the receiver. The Event field carries the value 1 when the member joins the group. However, the value 0 notifies the member departure. Furthermore, the AP notifies the appropriate receiver following any join or leave event. Therefore, BNAK defines a management packet, called Membership Notification, to inform the associated stations about their membership. This packet is shown in Fig. 2. The four first fields and the last field of the notification packet are standard elements. The Multicast Address field is used to identify the concerned session. The Starting Sequence Number field indicates the sequence from which the receiver is allowed to request packets. This field is required when a member joins a group in the middle of the session. Therefore, it does not request packets which were transmitted and still buffered for older members. The membership status is 1 if the receiver joins the group and 0 if it leaves the session. The Lowest Rate field indicates the minimum transmission rate that the AP can afford for the multicast session. The default PHY layer is OFDM since it is the one used in high throughput networks. Thus, the Lowest Rate field is encoded using 4 bits according to Table 18.6 of (Wireless LAN, 2012a). However, any other PHY layer may be used and signaled by allocating the reserved bits preceding the Lowest Rate field. The PER Limit field contains the value of the Packet Error Rate that the AP tolerates. This value is set on a basis of 10000, e.g., PER = 100 indicates a PER of 1%. The value of the PER Limit field varies between 1 and 10000. It allows a member to determine the coverage area of the multicast service. This coverage is infinite when PER = 10000. Note that the nil value as well as values higher than 10000 are reserved.

A Membership notification packet is acknowledged and, therefore, is transmitted reliably to the receivers.

### 3.2. Transmission procedure

The IEEE 802.11 defines robust modulation schemes, an efficient signal processing and powerful error correction codes. Thus errors are mainly due to collisions or inappropriate transmission data rate. In our previous study (Daldoul et al., 2015), we showed using a testbed that the device unavailability (i.e., configuration issues related to the receiver) may lead to an important loss rate. Then we showed that a receiver with suitable configuration experiences a very limited Packet Error Rate (PER) in a typical network under the following assumptions: (1) collisions are avoided, (2) the receiver is within the range of the sender, and (3) packets are transmitted using the appropriate PHY data rate. In order to ensure this limited PER, it is essential to select the appropriate data rate and to protect multicast packets against collisions. This protection may be guaranteed using standard features like CTS-to-Self or channel access during Contention Free periods. It may be achieved as well using Busy Symbol that we have proposed in (Daldoul et al., 2013).

The design of BNAK relies on the principle that the PER of the wireless network becomes very limited if the main loss factors (device unavailability, collisions and path-loss) are removed. Therefore, BNAK requires the use of suitable devices and a collision prevention feature, and it defines a retirement/reactivation procedure. In the remainder, we consider the use of CTS-to-Self to avoid the collisions. This feature is defined by the standard and is supported by all the existing devices. Besides, the retirement/reactivation procedure intends to protect the multicast session against members moving away from the sender. Such members will experience bad channel conditions due to the path-loss. Therefore, a receiver should retire temporarily (i.e., stop asking for retransmissions) when it is located beyond the suitable streaming area, and is allowed to reactivate when it comes back to this area. Therefore, only members located within appropriate distances will take advantage of a reliable multicast streaming. Once the principal loss factors are removed, it is more appropriate for the AP to request negative feedbacks from the multicast members experiencing losses, than requesting repeatedly BACKs for packets which are delivered correctly almost all the time.

Using the BNAK policy, packets are transmitted in blocks followed by a Block NAK Request (BNR). Only users encountering losses are invited to send a feedback. Therefore if all packets are transmitted correctly, no BNAK is transmitted and the bandwidth is saved. If a failure occurs, only the impacted receivers are allowed to send a feedback. In order to avoid eventual collisions between multiple BNAKs, these packets are transmitted after channel contention and are acknowledged by the AP. In other words, a BNAK is a unicast packet which is retransmitted if it is lost. Once the BNR is transmitted, the AP should contend for the channel before transmitting any other packet. It is hence possible that the AP gains the channel and transmits a new block before receiving the BNAKs, if any. An example of the BNAK procedure is provided in Fig. 3.

To summarize, BNAK protocol defines and uses 2 control packets to allow the feedback recovery:

- **BNR:** this control packet is transmitted by the AP and does not
need to be acknowledged by any receiver. It notifies the members about the delivered multicast packets. Using BNR, any member is able to detect and to request missing packets, if any. BNR is sent at the end of a block and is therefore protected against collisions.

- **BNAK**: this control packet allows the members to request the missing packets. BNAK is a unicast packet, so the AP should acknowledge it. If it is not acknowledged, the member retransmits it subject to the retry limit. As such, BNAK is delivered reliably to the AP. Besides, a member sends a BNAK after channel contention in order to avoid the collisions. We note that a member builds a BNAK only if it detects any missing packet. If all the multicast packets are delivered correctly to all the members, no BNAK is built and the bandwidth is saved.

The reliability of our protocol is based on the successful reception of the BNR. But this packet may be lost. To avoid the loss of the BNR, we transmit it using a robust data rate (or even the most robust, but the lowest one). Even though this packet is lost (this should be very exceptional), a multicast member builds its BNAK, if any, using the following BNR. We note that the loss of a BNR does not affect the reliability of our protocol if the multicast packets are received correctly.

It is worth noting that the use of a low data rate to send BNRs does not require much more transmission time compared to a high PHY data rate. This is because an important part of the transmission time of the packet is used by the SIFS plus the required time to transmit the PHY preamble and header. Table 1 illustrates the transmission time of the packet is used by the SIFS plus the required time to deliver the multicast packets. The default PHY layer is the OFDM one. Thus the Rate field is encoded on 4 bits according to Table 18.6 of (Wireless LAN, 2012a). However any other PHY layer may be used and signaled by allocating the reserved bits preceding the Rate field.

<table>
<thead>
<tr>
<th>Rates (Mbps)</th>
<th>BNR (25Bytes)</th>
<th>Data packet (1500B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>76 μs</td>
<td>2040 μs</td>
</tr>
<tr>
<td>12</td>
<td>56 μs</td>
<td>1040 μs</td>
</tr>
<tr>
<td>24</td>
<td>48 μs</td>
<td>540 μs</td>
</tr>
<tr>
<td>48</td>
<td>44 μs</td>
<td>288 μs</td>
</tr>
<tr>
<td>54</td>
<td>44 μs</td>
<td>260 μs</td>
</tr>
<tr>
<td>130</td>
<td>48 μs ; 56 μs</td>
<td>140 μs ; 148 μs</td>
</tr>
</tbody>
</table>

It is necessary to define robust actions in order to avoid useless BNAK transmissions. Also the member should not send a BNAK which requests missing packets in addition to packets being received correctly. Therefore we define the following 3 different statuses of a packet:

- **OK**: the packet is correctly received and does not need any retransmission;
– Missing: the packet is lost and either (1) the receiver sent a BNAK and is waiting for the retransmission of the missing packet or (2) the receiver is waiting a BNR in order to build a BNAK;

– Pending: the member generated a BNAK to request the retransmission of this packet. However, the BNAK is still buffered or being delivered, but the delivery has not finished yet (i.e. ACK not received yet). When the packet status is Pending, the member should not build a new BNAK to request the pending packet, unless the status of at least one other packet is Missing. In this case, upon the reception of a BNR, the member should (1) delete the pending BNAK, (2) modify the status of Pending packets to Missing, (3) build a new BNAK and (4) set the status of all missing packets to Pending. Upon the final transmission success or failure of the BNAK (failure due to reaching the retry limit), the status of all pending packets is set back to Missing.

A member may receive correctly a packet having the status Pending (the packet is retransmitted based on the feedback of another member which experienced the same failure and gained access to the channel first). In this case this receiver (1) sets the status of the corresponding packet to OK, (2) deletes the buffered BNAK (not adequate anymore), (3) sets any pending packet to Missing and (4) waits for the reception of a new BNR in order to build a new BNAK, if still required.

It is possible that the AP receives BNAKs requesting the retransmission of more packets than what the AP is allowed to send within a block. In this case the AP retransmits as many packets as possible within a TXOP, and builds a BNR which has in the Last Sequence field the sequence number of the packet preceding the first missing packet which will be retransmitted in the next block. Therefore, the AP avoids any BNAK which may request a packet that the AP is willing to retransmit. As an example, suppose that the AP is allowed to send one single packet per block, but receives a BNAK requesting packets 1 and 4. Then the AP retransmits packet 1 followed by a BNR having 3 in the Last Sequence field and contends again for the channel to send the second missing packet. Hence the AP avoids BNAKs requesting packet 4 before the retransmission of this packet.

On the other hand, if a member receives a BNR and finds that the sequence number of at least one pending packet is no longer in the BNR window, in this case this member should delete the pending BNAK and build a new BNAK, if still required, based on the new BNR.

We highlight that all 802.11 devices deliver a copy of all transmitted packets to their drivers and notify about the transmission success or failure. This notification allows a member to update the status of Pending packets following the transmission of a BNAK. Fig. 5 depicts the sequence diagram with transitions between different statuses.

We consider the special scenario illustrated in Fig. 6 in order to clearly introduce the required actions for a robust protection against useless BNAK transmissions and useless data retransmissions. We intentionally omit the collision protection feature in this figure for a better quality of the presentation.

At \( t = t1 \), members M1 and M2 receive packets 1 and 3 successfully. However they fail to receive the same packet having the sequence number 2. Therefore they record the following information:

\[ M1(t1): 1 \rightarrow OK; 2 \rightarrow Missing; 3 \rightarrow OK, M2(t1): 1 \rightarrow OK; 2 \rightarrow Missing; 3 \rightarrow OK. \]

At \( t = t2 \), M1 and M2 receive a BNR. Thus each of them builds a BNAK to request packet 2. Once the BNAK is buffered, the two members record the following information:

\[ M1(t2): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK, M2(t2): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK. \]

At \( t = t2 \), M1 and M2 receive a BNR. Thus each of them builds a BNAK to request packet 2. Once the BNAK is buffered, the two members record the following information:

\[ M1(t2): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK, M2(t2): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK. \]

Hence, M1 and M2 will not generate a new BNAK upon the reception of a new BNR. At \( t = t3 \), M1 sends a BNAK but the AP misses the feedback (for example, due to a collision). Since this is not the last transmission attempt of BNAK, M1 does not change the status of packet 2. At \( t = t4 \), M1 receives correctly packets 4 and 5 while M2 misses a new packet. Hence their respective records become as follows:

\[ M1(t4): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK; 4 \rightarrow OK; 5 \rightarrow OK, M2(t4): 1 \rightarrow OK; 2 \rightarrow Pending; 3 \rightarrow OK; 4 \rightarrow OK; 5 \rightarrow Missing. \]

Therefore, when M1 receives a new BNR at \( t = t5 \), it does not perform any action. Note that if there are only two statuses: OK

![Fig. 4. BNR and BNAK packet format.](Image)

![Fig. 5. Packet status sequence diagram.](Image)
and Missing, in this case M1 builds a new BNAK at t=t5 to request packet number 2, even though a BNAK is already in the queue. Therefore, the status Pending is necessary in order to avoid useless transmissions. On the other hand, M2 deletes its pending BNAK which does not include the new missing packet. Then M2 sets the status of packet 2 to Missing, generates a new BNAK for packets 2 and 5, and sets the status of these two packets to Pending. Thus we obtain the following records.

M1(t5): 1→OK; 2→Pending; 3→OK; 4→OK; 5→OK.
M2(t5): 1→OK; 2→Pending; 3→OK; 4→OK; 5→Pending.

At t=t6, M2 concludes the success of the BNAK transmission. Thus it sets the status of packets 2 and 5 to Missing. The records of M1 remain unchanged. At t=t7, M1 and M2 receive packet 2 successfully. Thus their records become as follows:

M1(t7): 1→OK; 2→OK; 3→OK; 4→OK; 5→OK.
M2(t7): 1→OK; 2→OK; 3→OK; 4→OK; 5→Missing.

Therefore M1 deletes its buffered BNAK which perished due to the successful reception of the pending packet. At t=t8, M2 receives packet 5 correctly and sets its status to OK. Thus at t=t9, when M1 and M2 receive a BNR, they do not build any BNAK since they have received all the multicast packets correctly.

3.4. Member retirement and reactivation

The data rate adaptation is another requirement to deal with the multi-rate capability of 802.11 WLANs. However, when an adaptation algorithm is used together with BNAK, the AP may select the lowest rate in order to establish a reliable communication with the farthest group member. This reduces the overall network throughput significantly and increases the congestion probability. We note that the application-level bit rate of multicast flows is not adjustable. Thus the AP may drop many packets during the congestion periods. These packets will not be retransmitted and will be lost definitely. When the network is saturated, it leads either to a significant distortion of the user QoE or to the total service interruption. Besides, the quality deterioration will be experienced by all the group members even though supporting high transmission rates. Therefore, the other alternative to the dynamic rate adaptation approach is to select a transmission rate statically. This solution ensures the use of a transmission rate which satisfies the throughput requirements of most exigent services like the video streaming application. Moreover, it prevents farther members from disturbing the multicast session. In this paper we consider the static selection of the transmission rate. The rate adaptation scheme will be investigated in our future works.

For a given multicast session, the AP notifies the group members about the lowest allowed PHY data rate. This notification is achieved using the BNAK Notification Message, whenever a new member joins the session. Several methods may be considered to configure the lowest rate per multicast address. The easiest one is to set one single rate to all the sessions. This is the default configuration of current AP where the lowest supported data rate is used. Another method is to offer a configuration tool (e.g. a local web interface) to the network administrator to set the lowest allowed rate per session. However, we recommend the updating of IGMP/MLD in order to include the bandwidth requirements of the multicast service. This information is then communicated to the MAC layer of the AP using internal messages, as previously described. Based on the required application bitrate, the AP selects a transmission rate which avoids the congestion and ensures the real-time streaming. We note that the way the transmission rate per multicast address is configured is beyond the scope of this work.

When the multicast packets are delivered using a transmission rate higher than the lowest one supported by the physical layer, a member may miss all the transmitted packets if it is located out of the coverage area of the used rate but within the range of the AP. However, this member remains able to detect these losses. This is because the BNR is sent at the lowest PHY rate, and is received correctly regardless of the location of the associated station. Thus, it is necessary to prevent this receiver to send BNAKs in order not to disturb the streaming session. We highlight that it is useless to retransmit the missing packets because the member is already beyond the coverage area of the multicast session. Therefore, a receiver should retire temporarily until it comes back to the streaming area. We note that the retirement and the reactivation decisions are achieved internally by the member and without the need for any request or packet exchange with the AP.

3.4.1. Member retirement

A member should retire if the SNR corresponds to a loss rate exceeding PER Limit at Lowest Rate. These two values are communicated to the members using the BNAK Membership Notification packet each time a new receiver joins the session. We note that a member may join the group while it is located out of the coverage area of the multicast service. In this case this member retires immediately upon the reception of the membership notification. Following the reception of a BNR, a retired member should set all missing packets as received correctly. Thus, it does not request them when reactivating.

3.4.2. Member reactivation

When a member retires, it should wait until its connection link improves before reactivating. This improvement is based on the SNR increase. The reception signal strength is time-averaged based on recent packets received correctly from the AP, and particularly following the successful reception of BNR, multicast packets and beacons. The SNR is then mapped to bit error rate. Since the PER depends on both the bit error rate and the packet length, we measure the packet loss rate on the basis of 1538 byte-length packets. This size corresponds to the maximum size of Ethernet frames in addition to the MAC header and is the typical size of packets belonging to a video stream. We note that SNR is easily mapped to PER (Vutukuru et al., 2009; Halperin et al., 2010; Rayanchu et al., 2008) However, this mapping may vary slightly from one chipset to another. This is because some high quality devices have enhanced sensitivity compared to other receivers. Therefore we recommend the implementation of a vendor-specific mapping function.
We define PER Low as the indicator of the eligibility of a retired member to reactivate. This value is lower than PER Limit. A member reactivates when the PER is lower than PER Low. We use different PER values to retire and to reactivate in order to avoid frequent retirement and reactivation when the receiver is at the edge of the multicast service area. To determine the relationship between PER Limit and PER Low, we illustrate the PER at different distances using different data rates in Fig. 7.

The results of Fig. 7 are obtained using NS-3 according to the simulator configuration of Table 3. We observe that these curves are almost linear. Hence a ratio of 100 (i.e. \( \text{PER Low} = \frac{\text{PER Limit}}{100} \)) implies a reactivation sub-area of about 90% of the entire multicast area, regardless of the transmission rate and PER Limit.

Unlike PER Limit, the value of PER Low is not defined by the AP and is configured by the receiver itself. In the remainder of this paper, we measure the PER using the Nist-Error-Rate model of NS-3.

During the retirement period, the receiver remains member of the group but stops sending BNAKs. This is because the loss rate exceeds the allowed limit. However, the viewer may ignore the reason of the video quality deterioration or the total streaming interruption. Therefore we recommend that the MAC layer sends a notification to the video player or to the operating system in order to report a temporal disruption due to weak signal strength. An alert message is then displayed at the front of the user’s screen. Examples of retirement notifications are illustrated in Fig. 8. This motivates the client to move toward the AP, and enables a quick reactivation of the MAC layer.

### 3.5. Loss factors management in WLAN networks

Packet losses in a wireless network may occur for several reasons. The principal loss factors are the following: interferences, collisions, path-loss, hidden station, Doppler Effect and device unavailability. They are described like follows:

- **Interferences**: There are two kinds of interferences, (1) those which occur between different 802.11 networks which use overlapped channels, and (2) interferences between a 802.11 network and another device using the same frequency, such as wireless phones and Bluetooth devices. The first kind is generally easy to fix thanks to a good planning of the radio band. This is possible as there are an important number of non-overlapped channels at different bands: 2.4 GHz, 5 GHz, 60 GHz, etc. The second kind of interferences is real because WLANs use unlicensed frequencies that can be used by any other technology. However, regulators impose limits for the transmission power over these frequencies. So the WLAN and the interferer should be located within few meters from each other for the interference to occur. In many cases, the interfering devices belong to the WLAN owner, and may be disabled to enhance the WLAN performance. Another solution is to avoid the 2.4 GHz band which is mostly used by interfering sources. So the 5 GHz band may be a good option.

- **Collisions**: They occur when two stations start their transmissions simultaneously. The standard defines the Backoff procedure and an exponentially increased Contention Window (CW) to reduce the collision probability. This solution cannot completely eliminate the collisions. We note that the collision rate increases when the network load is high. This is because stations will have always data to send and will contend for the channel frequently. But if the network load is low, collisions will be limited.

- **Path-loss**: As the distance between the sender and the receiver increases, the signal reception strength decreases. The reception quality is measured as the Signal to Noise Ratio (SNR). To deal with path-loss, the standard defines many data rates where high data rates are useful for high SNR values, and low data rates are efficient with lower SNR.

- **Hidden station**: A station may be hidden from station “A” but within the range of “B”. So it is not able to hear transmissions from “A” to “B” and may start a transmission while “A” is sending data to “B”. The hidden node problem is mainly present in ad hoc and mesh networks. In these networks, devices are deployed.

![Fig. 7. Packet Error Rate as a function of distance for 1538 byte-length packets; use of the Nist-Error-Rate model.](image7)

![Fig. 8. High level notification about the MAC layer retirement.](image8)
randomly without any planning of the radio resources. According to the standard, hidden nodes may also exist in the infrastructure mode if nodes use very low transmission power. However, the AP is the central node in the infrastructure mode and is not hidden to any associated station. So the AP’s transmissions are not affected by the hidden node problem. Besides, the AP is the only station allowed to use multicast transmissions (other stations should send any multicast traffic using unicast transmissions to the AP). So multicast is not affected by the hidden station in a typical infrastructure network.

- **Doppler Effect**: This occurs when nodes move at high speeds (Zhipeng, 2009). So Doppler Effect is not a problem for WLANs where users are expected to be static or to move at human speeds.

- **Device unavailability**: We show in a previous study (Daldoul et al., 2015) that a device with a bad configuration or a highly loaded CPU may be unable to receive packets. We conclude that the use of an appropriate device is able to reduce the experienced loss rate.

Some other loss factors may exist, such as fading and multipath effect. However, the standard defines robust modulation schemes, an efficient signal processing and powerful error correction codes to combat them. Thus, these additional loss factors should incur a low loss rate in a typical WLAN deployment, but they may incur more losses in specific industrial environments such as (Willig et al., 2002). We note that if the loss rate of a WLAN is high, even the unicast becomes unreliable and the network performance may be very limited. This is because the default unicast acknowledgment policy allows a limited number of retransitions upon transmission failures. If we consider the case of ath9k driver (ath9k), a unicast packet will be definitely lost if its 4th retransmission fails. In this case, a packet loss rate of $p = 50\%$ leads to a unicast loss rate of $pu = p^5 = 6.25\%$. To summarize, the impact of a high loss rate in a WLAN are the following:

- The unicast becomes unreliable. If it is used to deliver a UDP stream, the receiver will experience packet losses. But if the unicast is used to send a TCP flow, the TCP source will retransmit the missing packets but will reduce its data rate (Floyd et al.).
- Many packets are transmitted several times, wasting to available transmission time.
- Frequent use of large Contention Windows (CW), because the sender should increase its CW upon a transmission failure (Wireless LAN, 2012a). This increases the average waiting time for the channel access and limits the network throughput.
- Frequent use of low data rates, as most rate adaptation algorithms (Pefkianakis et al., 2011; Acharya et al., 2010; Huang et al., 2013) switch to a lower data rate upon transmission failures.

All of these consequences lead to a significant degradation of the WLAN performance, and 802.11 networks may even be useless if their loss rate is high. Such networks are useful and performant in typical environments when the loss rate is limited. For such environments we define our protocol, BNAK.

### 3.6. BNAK applicability

We define BNAK to deliver multicast flow over WLANs in the infrastructure mode. Our protocol is (1) reliable: it uses a feedback policy to detect and retransmit missing packets, (2) scalable: it is able to support an important number of receivers without or with very limited impact on the throughput, (3) efficient: it needs limited overhead and can take advantage of recent features such as packet aggregation (A-MSDU and A-MPDU), and 4) standard compliant: BNAK uses the standard channel access procedure and does not need any modification of the multicast packet format.

To take a full advantage of BNAK, the network loss rate should be limited. Providing a multicast service for a large group is generally required by network providers in professional WLAN deployments (e.g. airport, hotel, stadium, company, etc.). A network administrator who wishes to provide a reliable multicast flow for a large multicast group should make a good planning of the radio spectrum in order to get rid of the interferences. Besides, as our protocol is defined for the infrastructure mode where the AP is the only multicast sender, it is clear that BNAK is not concerned with the hidden station problem. Regarding the path-loss, a possible solution is to use a rate adaptation algorithm. The definition of such an algorithm is beyond the scope of this work. However, we introduce another solution for the path-loss which is “the member retirement and reactivation procedure”: a member should stop requesting retransmissions whenever it is located beyond the multicast streaming area.

Besides, we consider that the group members use suitable devices and are static or move at human speeds. So we get rid of device unavailability and Doppler Effect. The next loss factor is collisions. We note that our protocol does not count on the random Backoff to eliminate the collisions. Instead, BNAK counts on CTS-to-Self transmitted at the highest data rate. The standard defines this frame to reserve the channel. However, we use it for another purpose which is “the collision protection”. We choose to send it at the highest data rate to reduce its transmission time and to minimize the overhead. The correct reception of this frame by the multicast members is not required for the good operation of our protocol. So the AP sends a CTS-to-Self before delivering any multicast packet. After the transmission end of the CTS-to-Self, the AP senses the channel again. If the medium is busy, this means that a collision is happening, so the multicast transmission is canceled and will be scheduled in the next channel access. Otherwise, the AP sends the multicast packets separated with “SIFS”. The SIFS period is short and ensures that no contention for the channel may occur during the transmission. So the packets are delivered without any collision. Fig. 9 depicts the protection using CTS-to-Self in the absence and in the presence of a collision.

Once the principal loss factors are removed, the remaining loss rate in a typical network will be limited. So BNAK will recover any missing packet and will achieve very high performance as will be demonstrated further in this paper.

![Fig. 9. BNAK and multicast protection against collisions using CTS-to-Self.](image-url)
We note that a good planning of the radio spectrum may be a complicated task for a personal deployment of a network. However, in this deployment scenario, multicast may be needed for small groups of about 2 or 3 members. So the AP may use an algorithm to select the most suitable multicast protocol among supported ones.

To conclude, BNAK is a multicast protocol defined for the infrastructure mode where the AP is the only multicast sender. Our protocol is reliable, efficient, scalable and standard compliant. It allows the network provider to server large multicast groups reliably and efficiently. However, to take a full advantage of any WLAN and of BNAK as well, it is necessary to make a good planning of the radio spectrum. This is because randomly deployed networks may experience low performances and are not always able to ensure enough bandwidth for high throughput streams even those using unicast.

4. Model description

In this section we define an analytical model to evaluate the efficiency, i.e. throughput, and the scalability of BNAK. We consider multicast UDP/IP packets with the maximum transmission unit (i.e. 1500 Bytes). Thus the MAC packet length is 1538 Bytes. We consider that BNARs are always transmitted successfully.

The key approximation of our model is that BNAKs do not collide and are always transmitted correctly to the AP. This is a valid approximation when the PER is limited since BNAKs are supposed to be non-frequent and the channel contention is very limited. Therefore our model provides a high accuracy when the PER is limited.

Let G be the multicast group size. Each member in the group experiences a PER of pi for i = 1..G. We consider that losses are not correlated between different receivers. We fix the transmission limit of a packet to 100. We choose this value because the retransmission of a packet is subject to lifetime limit. Thus we fix a transmission limit by excess. We note that the probability to reach high transmission stages is negligible when the PER is limited. Let Nr be the block size and Nr(k) be the number of packets transmitted for the kth time within a block. Nr(1) is the number of packets transmitted for the first time. Every block is composed of \( \sum_{k=1}^{100} Nr(k) \) packets. Nr(k), k = 1..100, depends on the PER of the network. Table 2 presents the used variables and their values at different transmission data rates. We consider that BNR, BNAK and ACK are always transmitted at 6 Mbps. On the other hand, the CTS-to-Self is always delivered at the highest data rate of 54 Mbps in order to have the shortest length. Thus it allows the efficient detection of simultaneous transmissions.

We define X as the number of transmission attempts. The probability for a given member Mi, i = 1..G, to receive correctly a packet in any of the k first transmissions is given by:

\[ P(X \leq k) = 1 - p_i^k \]

We derive the probability to serve all the G receivers in any of the k first transmissions as follows:

\[ P(G \leq k) = \prod_{i=1}^{G} (1 - p_i^k) \]

We obtain Nr(1) and Nr(k), k = 2..100, using Eqs. (3) and (4) respectively.

\[ Nr(1) = \sum_{k=1}^{100} Nr(k) \]

\[ Nr(k) = N(1) \cdot (1 - P(G \leq k-1)), k = 2..100 \]

We resolve Eqs. (3) and (4) and we obtain Nr(k), for k = 1..100, as shown in Eq. (5). It is obvious that the probability to receive a packet correctly in X = 0 attempt is nil, hence P(X = 0) = PG(X = 0) = 0.

\[ Nr(k) = \frac{N(1) \cdot (1 - P(G \leq k-1))}{\sum_{i=1}^{G} (1 - P(G \leq k-1))}, k = 1..100 \]

In Eq. (6) we derive the probability to deliver correctly a block to a given member Mi, i = 1..G. This equation allows us to compute the probability of a BNAK generation.

\[ P(N) = \prod_{k=1}^{100} (1 - p_i^k) \]

We compute the average deferral time as follows. Suppose nodes n1, n2 and n3 contend for the channel. Using our no-collision approximation, the 3 nodes generate different Backoff Times (BT). For example: 2 for n1, 5 for n2 and 7 for n3. Thus n1 transmits first, after DIFS + 2 × SlotTimes at the end of the current transmission. Then n2 transmits, DIFS + 3 × SlotTimes after the transmission end of n1. Finally n3 transmits, DIFS + 2 × SlotTimes after the transmission end of n2. Thus the total deferral time in this example is 3 × DIFS + 7 × SlotTimes, and the maximum deferral time which may occur is 3 × DIFS + CWmin × SlotTime. As we suppose that BNAKs are not frequent, we consider that the average Backoff Time is BT = CWmin × 2. Therefore, the average deferral time for G = 1 contending nodes, is: DIFS + CWmin/2 × SlotTime + G0 × DIFS. In Eq. (7) we obtain the average packet transmission time using the BNAK policy.

Table 2

<table>
<thead>
<tr>
<th>Variables description and value.</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>IEEE 802.11a</td>
</tr>
<tr>
<td>TPPDU_Data: PHY packet duration, 1538 B.</td>
<td>232 μs (at 54 Mbps)</td>
</tr>
<tr>
<td>TPPDU_BNR: PHY BNR duration, 25 Bytes</td>
<td>60 μs (at 6 Mbps)</td>
</tr>
<tr>
<td>TPPDU_BNK: PHY BNR duration, 30 B.</td>
<td>64 μs (at 6 Mbps)</td>
</tr>
<tr>
<td>TPPDU_ACK: PHY ACK duration, 14 B.</td>
<td>44 μs (at 6 Mbps)</td>
</tr>
<tr>
<td>PROTECTION_DURATICTION, CTS</td>
<td>40 μs (at 54 Mb + SIFS)</td>
</tr>
<tr>
<td>PROTECTION_DURATICTION, BS</td>
<td>16 μs (SIFS)</td>
</tr>
<tr>
<td>SlotTime</td>
<td>9 μs</td>
</tr>
<tr>
<td>SIFS</td>
<td>16 μs</td>
</tr>
<tr>
<td>DIFS (SIFS + 2 SlotTime)</td>
<td>34 μs</td>
</tr>
<tr>
<td>CWmin: Contention Window min</td>
<td>15</td>
</tr>
</tbody>
</table>
\[ T_{\text{BNAK}}(N, G) = \frac{\text{DIFS} + \frac{\text{CW}_{\text{min}}}{2} \times \text{SlotTime} + \text{PROTECTION}}{\text{DURATION}} + (T_{\text{PPDU\_Data}} + \text{SIFS}) \times N + T_{\text{PPDU\_BNAK}} + \left( \sum_{i=1}^{G} (1 - P(N)) \right) \times (\text{DIFS} + T_{\text{PPDU\_BNAK}} + \text{SIFS} + T_{\text{PPDU\_ACK}}) / Nr(1) \] (7)

Eq. (8) gives the packet transmission rate of BNAK.

\[ \text{Throughput}_{\text{BNAK}} = \frac{1}{T_{\text{BNAK}}(N, G)} \] (8)

5. Simulation results

We use NS-3 to validate our analytical model and to evaluate the performance of our protocol. We build an IEEE 802.11a infrastructure network and we consider the simulation parameters in Table 3. We consider multicast packets of 1538 Bytes (including the MAC header) transmitted at the highest rate of 54 Mbps. Besides, the CTS-to-Self is continuously sent at 54 Mbps while the other control packets (i.e. BNR, BNAK, ACK, BAR, BACK) are always delivered at the lowest rate of 6 Mbps.

5.1. Model validation

To validate our analytical model we consider that all the group members have the same PER. We compare the analytical and simulation results of BNAK in Fig. 10. We consider groups of 1, 10 and 100 members. We notice that there is a good match between the simulator and the mathematical model for limited values of PER. As expected, we observe a small gap between the analytical and the simulation results when both the loss rate and the group size increase. Particularly we find that the analytical results exceed the simulation ones when the group size is 100 and the packet error rate varies between 0.1% and 3%. This gap is caused by the collisions between the BNAKs which are not considered by the analytical model but impact the throughput in the simulations.

However, the simulation throughput becomes higher when the packet error rate is between 20% and 50%. This is because a few BNAKs are enough to request the retransmission of all the missing packets. Therefore several feedbacks are canceled under the simulator. However, the analytical model does not consider these suppressions. This explains the reversed gap at high error rates.

Then we validate our model for different block sizes. We consider a multicast group of 10 members and two different PER values of 0.01% and 1%. We compare the analytical and the simulated throughput. Fig. 11 illustrates the obtained results. We observe that our model has a high accuracy for all the considered block sizes. Therefore it can be used to determine the throughput based on the block size.

5.2. Performance analysis

We evaluate the scalability of BNAK compared to GCR-BACK, GCR-UR2 (i.e. each packet is transmitted twice using UR policy) and the legacy multicast procedure. We set all the group members at a distance of 10 meters from the AP and we transmit the flow at 54 Mbps. We consider that no traffic other than the multicast stream is transmitted. Besides, we consider that the AP is in the saturation condition. We illustrate the obtained results in Fig. 12 for variable sizes of multicast groups. These results show that the scalability of BNAK is unlimited when the receivers are located at an appropriate distance from the sender. We observe that the throughput of GCR-BACK decreases significantly when the group size increases. Particularly when 100 members are in the group, we notice that GCR-BACK delivers only 268 pps while our protocol achieves more than 3250 pps. Besides, these results demonstrate that the redundant retransmissions limit the efficiency of the Unsolicited Retry policy significantly. Hence, the throughput of GCR-UR2 is less than 50% of that of BNAK. Moreover, we find that our protocol outperforms the legacy transmission procedure. This is because BNAK takes advantage of the high efficiency of the block transfer, while the conventional multicast delivers each packet following channel contention and waiting periods.

In Figs. 13(a) and 14(a) we consider a group of 10 members. We observe that BNAK outperforms 802.11aa significantly and provides more than twice the throughput of GCR-BACK when the loss rate is limited. Moreover BNAK outperforms the legacy multicast at distances below 25 m. When the loss rate increases, we observe that the legacy multicast provides higher throughput on the
expense of limited delivery ratio.

Figs. 13(b) and 14(b) illustrate the throughput and the reliability, respectively, for a group of 100 members. We observe that the throughput and the reliability of the legacy multicast do not depend on the group size. At the same time we notice that the throughput of BNAK is not impacted by the group size when the receivers are within the coverage area of the sender. Moreover we observe that BNAK delivers about 12 times the throughput of GCR-BACK at reasonable distances from the AP.

We consider two group sizes of 10 and 100 members and we evaluate the throughput and the reliability of these protocols under a variable loss rate. We consider that all the members are located at the same distance from the AP in order to ensure the same PER for all the receivers. We consider the saturation condition in order to measure the highest throughput. Thus the transmission queue of the AP is never empty. We consider blocks of 5 packets.

In Figs. 13 and 14 we consider the case of an unshared network. Hence the AP is the only sender and the medium is entirely used by the multicast traffic. Therefore, the reliability of the legacy multicast is not reduced by the collisions.

Then we consider a medium shared between the multicast traffic of the AP and a unicast flow generated by an associated station. The unicast traffic is transmitted at 54 Mbps using the
basic feedback policy (i.e. individual transmission and acknowledgement). The unicast sender is in the saturation condition and its transmission queue is never empty. This scenario may occur in a WLAN when a client uploads a large data file while a multicast session is ongoing.

Fig. 15 depicts the delivery ratio for groups of 10 and 100 members. We notice that the reliability of the legacy multicast is impacted by the collisions and that the delivery ratio of the standard procedure is about 90% when the receivers are at reasonable distances from the sender. However, we observe that the reliability of BNAK, GCR-BACK and GCR-UR is ensured thanks to CTS protection.

We vary the block size and measure the throughput of BNAK and GCR-BACK. We consider groups of 10 and 100 members. We install the multicast receivers at the same distance of 10 m from the AP in order to obtain the same loss rate. Fig. 16 depicts the obtained results. We observe that BNAK has a high efficiency even if it is used to deliver one single packet per transmission opportunity. Thus our protocol is also appropriate for low throughput streams such as voice and audio. On the other hand we observe that the throughput of 802.11aa for groups of 10 members is 472 pps when multicast packets are acknowledged individually. However, this protocol is not appropriate for large groups even for delivering low throughput flows.

Fig. 16. Throughput versus block size.

We evaluate the delays as a function of the throughput. The multicast sender generates packets at equal intervals and at a constant rate, and we increase this rate progressively. All the group members are located within a distance of 10 meters from the AP. We depict the obtained results for groups of 10 and 100 members in Fig. 17(a) and (b), respectively. As expected we observe that the incurred delays using BNAK do not depend on the group size since the experienced packet error rate is very low. Moreover, we notice that these delays increase when we exceed the capacity of our protocol which is 3251 pps. When the AP is in the saturation condition, the delivery latency is also impacted by the buffering delays and is bounded by the packet lifetime limit. We observe that the maximum delay of BNAK in the saturation condition is limited to 6.7 ms. This value depends on the queue size and the average service time. In our configuration of Table 3 we consider a maximum size of 20 packets. Thus, packets are dropped if they arrive while the queue is full. However, when a new packet is buffered, it should wait for the transmission of the first 19 packets.

On the other hand, the delays of 802.11aa BACK depend on the group size. According to Fig. 17(a), we notice that the delays of GCR-BACK increase slightly between 500 pps and 1500 pps. This is because a packet may arrive while BAR/BACK exchanges are in progress. In this case the new packet remains in the queue until the end of the feedback recovery phase before it is transmitted. Then, the delays increase again in the saturation condition. These delays are higher than those of our protocol because the average service time of GCR-BACK exceeds that of BNAK.

5.3. BNAK parameters analysis

In a WLAN, the main loss factors are collisions and path loss. Hence, BNAK uses a CTS-to-Self, and defines an activation/retirement procedure so that receivers with bad reception conditions do not disturb the multicast session. Therefore, the packet error rate is expected to be very limited. As BNR is sent at the end of the block, it is also protected against the collisions (the entire block is
protected with CTS-to-Self). Besides, it is sent at the lowest data rate. Thus, it is protected against the path loss (if a member fails to receive a BNR, it is likely that this member is out of the coverage area of the sender or is disconnected). However, exceptional BNR losses may occur. If this happens, then there are two possible scenarios. (1) BNR is missing but the multicast packets are received correctly: the missing BNR has no impact on the reliability/ delays of our protocol. (2) BNR is missing and at least one multicast packet is lost: the next BNR allows the receiver to detect and request the missing multicast packet. Thus, the second scenario does not affect the reliability of BNAK, but adds a temporal delay. This delay depends on 2 parameters. (1) The delay between 2 successive BNRs: our protocol is designed for real time videos that have a typical rate of 25 pictures per second (one image each 40 ms), thus at least one BNR is delivered each 40 ms. (2) The probability that scenario 2 occurs: the probability of losing BNR and another packet simultaneously is much lower than the probability of losing BNR only. Therefore, delays due to scenario 2 have no impact on BNAK because they are multiplied by a very small probability. Besides, their temporal effect is negligible for real time streaming applications which tolerate a latency of about 1 s.

On the other hand, BNAK packets are not protected against collisions, but they are acknowledged. Thus, if a BNAK is missing, the member sends it again. Besides, the collision probability is a function of the members having a BNAK to send (i.e. the number of receivers experiencing multicast packet losses). If all the stations of a group of 100 members receive the multicast packets correctly, none of them sends a BNAK. But if only one member among the group has missing packets, only this member sends BNAK. As our protocol relies on loss prevention, the loss rate is expected to be very low, and a very limited number of BNAKs are transmitted under ordinary operating conditions (i.e. receivers within the coverage area of the sender). So none or few BNAKs are sent after a BNR, even for large multicast groups. If a BNAK is lost, the member sends it again (recall that BNAK packets are acknowledged).

In order to provide a reliable protocol, reducing the latency to retransmit missing multicast packets is essential. Therefore, we evaluated the latency and shown the obtained results in Fig. 17. Also, we considered the overhead incurred by BNAKs in Section 5 (Simulation Results). Specifically, the overhead is shown in Figs. 10 and 13 when the packet error rate (or distance) increases for individual and block transfers, and supports the packet aggregation feature of 802.11n seamlessly.

6. Conclusion

In this paper we introduced BNAK, a new multicast protocol for 802.11 networks. We have also proposed a membership detection function for the MAC layer. This function runs at the AP and does not need any updates at the receivers. Then we described the packet format and the operating mode of our protocol. Furthermore, we modeled the retirement and the reactivation decisions of a moving member based on the reception SNR. We devised an analytical model to measure the throughput of BNAK under various configurations and different values of group size, block size and packet error rate. We validated the analytical model using simulations. The simulation results confirmed the accuracy of our model in measuring the expected throughput. They also showed that BNAK is reliable, and outperforms 802.11aa by a significant margin. In particular, the throughput of our protocol can exceed 10 times that of GCR-BACK.

The obtained results suggest that BNAK is reliable, efficient, and incurs limited transmission delays. Moreover, the scalability of BNAK is high when the causes of packet loss are eliminated hence its appropriateness for large multicast groups. Finally, BNAK is easy to implement in current devices since it requires software updates only and no need for any hardware modifications. It is appropriate

References

Christensen, M., Kimball, K., Solensky, F., 2006. Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches. RFC 4541, May.
He, Y., Yuan, R. Ma, X., Li, J., 2008. The IEEE 802.11 Power Saving Mechanism: An Experimental Study. In IEEE WCNC. Las Vegas, Nevada, USA.
Kuri, J., Kasera, Sheh Kamar, 1999. Reliable Multicast in Multi-access Wireless LANs. In IEEE INFOCOM. NY, USA.
NACK-Oriented Reliable Multicast (NORM) Transport Protocol, RFC 5740, November 2009.
Tanigawa, Y., Yasukawa, K., Yamaoka, K., 2010. Transparent Unicast Translation to Multicast in 802.11. In IEEE WCNC. Las Vegas, USA.
Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, 2012a. IEEE std 802.11.