Telecommunication Systems
https://doi.org/10.1007/s11235-024-01203-9

Evaluation of the IEEE 802.11aa group addressed service in VHT Wi-Fi networks

Blas Gómez[1] · Estefanía Coronado[1,2] · José Villalón[1] · Antonio Garrido[1]

Accepted: 16 July 2024

Abstract

Multimedia content represents a significant portion of the traffic in computer networks, and COVID-19 has only made this portion bigger, as it now represents an even more significant part of the traffic. This overhead can, however, be reduced when many users access the same content. In this context, Wi-Fi, which is the most popular Radio Access Technology, introduced the Group Addressed Transmission Service (GATS) with the amendment IEEE 802.11aa. GATS defines a set of policies aiming to make multicast traffic more robust and efficient. However, Wi-Fi is constantly evolving, and as it improves and greater bandwidths and data rates become available, it is necessary to reevaluate the behavior of mechanisms introduced in past amendments. This is also the case with GATS, whose policies have different behaviors and adapt better to different channel conditions. These policies have been evaluated in the past on High Throughput networks. Still, none of the evaluations provided insights into the behavior of GATS policies in Very-High Throughput (VHT) physical layers in a realistic manner. This is extremely relevant as a greater available bandwidth can impact the decisions of the GATS policy configuration. Thus, in this work, we present an evaluation of the IEEE 802.11aa amendment with a VHT physical layer in a realistic scenario that uses Minstrel as a rate adaptation algorithm simulated in NS-3.

Keywords Wi-Fi · Multicast · IEEE802.11 · VHT · GATS

1 Introduction

The surge in real-time multimedia traffic, such as VoIP and IPTV, has been aggravated by the COVID-19 pandemic and the increase in telecommuting and online and hybrid teaching. As a result, this surge has severely affected the wireless infrastructure of businesses, schools, and universities. Many students and employees use the same stream to connect to lectures, meetings, or seminars. In both of these scenarios, the most popular Radio Access Technology (RAT) is Wi-Fi. Multicast transmissions have become essential to prevent the same content from being sent in multiple unicast streams and saturating the Wi-Fi channel's capacity. However, the contention-based medium access of this RAT brings challenging conditions for the multicast transmissions, especially for those applications with strict performance requirements such as real-time video or audio.

The intricacy added by the multicast transmissions hampers the use of the rate adaptation algorithms typically used in unicast transmissions, such as Minstrel [1]. These algorithms use the feedback provided by the Acknowledgments (ACKs) in the unicast transmissions to decide the Modulation and Coding Scheme (MCS) that provides the best trade-off between the robustness of the signal and the data rate to minimize the loss of frames and achieves the highest possible throughput. This feedback provided by the ACKs is suppressed in multicast transmissions to avoid the feedback implosion effect. This makes the retransmission of lost frames impossible, as there is no way to know if a frame has been received. Moreover, adapting the data rate according to the channel quality is also not possible as the channel conditions are unknown. To maximize the number of multicast

✉ Blas Gómez
blas.gomez@uclm.es

Estefanía Coronado
estefania.coronado@uclm.es

José Villalón
josemiguel.villalon@uclm.es

Antonio Garrido
antonio.garrido@uclm.es

1 High-Performance Networks and Architectures, Universidad de Castilla-La Mancha, Albacete, Spain

2 I2CAT Foundation, Barcelona, Spain

Published online: 01 August 2024

stations receiving the transmissions correctly, the standard employs the most robust MCS available to ensure reliable frame delivery regardless of each station's perceived channel quality or the distance to the Access Point (AP). This comes at the cost of the need for more airtime to transmit multicast frames and, therefore, a higher channel occupancy.

In this regard, the IEEE 802.11aa amendment [2] was presented to overcome these problems by introducing the Group Addressed Transmission Service (GATS) whose main objective is to enhance multicast communications' reliability in Wireless Local Area Networks (WLANs) while ensuring compatibility with existing commercial devices. To achieve this, the amendment introduces a series of multicast transmission policies that aim to improve the overall reliability of multicast services. These policies are designed to adapt to various network conditions, ensuring optimal performance and correct operation of the multicast services.

The rest of the paper is organized as follows. Section 2 provides an overview of the related work. Section 3 gives the background on the IEEE 802.11aa amendment and VHT networks. Section 4 describes the evaluation of the GATS policies introduced in the amendment, and Sect. 5 presents the conclusions.

2 Related Work

Different evaluations of this amendment have been carried out [3–8]. However, most of them assess the amendment in settings which are now outdated. For instance, the authors of [3] present an analytical model that evaluates the performance of GATS, and their assessment is carried out using the OMNeT++ network simulator with the IEEE 802.11n standard. They conclude that the performance greatly depends on the status of the network. However, they do not include Enhanced Distributed Channel Access (EDCA) or dynamic MCS selection. In [4], they also present another evaluation that includes EDCA using IEEE 802.11n. However, it still does not use any rate selection algorithm, which greatly influences the performance of some of the GATS transmission policies. In the same work, they provide a set of guidelines for the use of GATS. The authors of [5] also evaluate this amendment using simulations in OPNET to assess GATS's scalability, delay, efficiency, and reliability. In [6], the authors experimentally evaluate GATS and provide its implementation for real hardware. Their evaluation is carried out using a real testbed with the IEEE 802.11g physical layer and both synthetic traffic and real video. Their results confirm that each GATS offers a specific trade-off that adapts best to different situations. This study also emphasizes the complexity of Block Acknowledgement (BACK). In [7], the authors also present an experimental evaluation on an IEEE 802.11g network of the Quality of Experience (QoE) of multicast video transmissions using GATS.

However, since the release of the IEEE 802.11aa amendment and the publication of the above evaluations, Wi-Fi has evolved enormously, especially with the advent of IEEE 802.11ac and the Very High Throughput (VHT) physical layer. Thanks to this progress, Wi-Fi networks can achieve much higher throughputs. A direct consequence of this is that the crossover points (i.e., when switching from one GATS policy to another is optimal) have changed as well, but this remains to be studied, as the only previous work [8] that evaluates the IEEE 802.11aa amendment on VHT networks does not use rate adaptation algorithms, which is one of the main characteristics of a realistic scenario and has a big influence on the decision of the GATS policies to be used.

Considering this, this paper's contribution is threefold:

(i) We review the main operational functionality of the GATS defined by the IEEE 802.11aa amendment on VHT Wi-Fi networks;
(ii) we implement GATS policies on the NS-3 network simulator, and our version of NS-3 with the implementation of the GATS is made publicly available; and
(iii) we conduct a realistic performance evaluation of the GATS policies with the IEEE 802.11aa amendment on a VHT network using Minstrel as a rate adaptation algorithm.

To the best of our knowledge, this is the first work that evaluates the performance of the IEEE 802.11aa amendment on VHT Wi-Fi networks with dynamic rate selection and offers usage recommendations for this kind of network.

3 Background

3.1 The IEEE 802.11aa Amendment

The limitations of multimedia streaming services are the focus of the IEEE 802.11aa amendment. It increases the efficiency and reliability of multicast traffic while maintaining the performance of the rest of the traffic. To achieve this, the amendment presents GATS, a novel method to overcome the low reliability of multicast services and increase transmission efficiency. GATS consists of two policies, namely Directed Multicast Service (DMS) and Group Cast with Retries (GCR), the latter being a set of three policies aimed at improving the reliability of multicast traffic and composed of: no retry/no ACK (NR/NACK), which is the traditional transmission policy that, in this work, we will refer to as Legacy, Unsolicited Retries (UR), and Block ACK (BACK). All the policies are described in detail below.

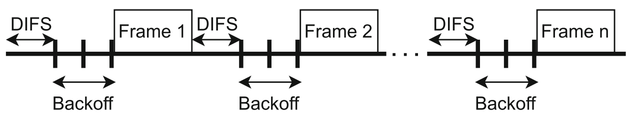


Fig. 1 Exchange of a set of n multicast frames using Legacy. The frames are addressed to the multicast group's address and can be received by all the stations that subscribed to the group

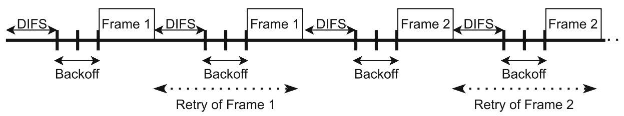


Fig. 2 Exchange of 2 multicast frames using GCR-UR policy. Each frame is retransmitted twice. All frames are addressed to the multicast group's address and can be received by all the stations that subscribed to the group

3.1.1 Legacy

The first one, which we will refer to as Legacy, is the traditional multicast mechanism introduced in the original IEEE 802.11 standard. When using this policy, frames are neither acknowledged nor retransmitted. In other words, the transmitter has no feedback on whether frames are received or not, as shown in Fig. 1. To ensure that the maximum number of Multicast Receivers (MRs) receive the multicast frames correctly, Legacy transmits at the most robust data rate, which allows it to reach MRs with a lower signal quality at the cost of a less efficient transmission. This results in Legacy frames occupying the channel for longer. However, certain vendors offer the flexibility to configure the transmission rate for Legacy transmissions.

3.1.2 GCR-UR

Using the basic rate makes frames less vulnerable to interference, but there is still no guarantee that frames will be delivered correctly to all the multicast group members. This is why GCR incorporates the UR transmission policy. In this case, as shown in Fig. 2, frames are always retransmitted, even if they were received on the first transmission (note that the transmitter cannot know this as frames are not acknowledged either when using UR). This policy exhibits lower reliability compared to DMS, but it solves the scalability issue by decoupling the reliability from the number of Multicast Receivers (MRs) in the network. However, many retransmissions might be unnecessary, resulting in a waste of resources.

3.1.3 GCR-BACK

This extends the BACK used for unicast frames to support multicast transmissions. In unicast, the sender and the receiver agree on transmitting a number of frames. The sender then requests the confirmation of all of them at once with a single ACK. In the case of multicast transmissions, all group members agree on using BACK. Then, the transmitter addresses the agreed-upon number of frames to the multicast group. After that, the transmitter sends individual BACK requests to each group member, as shown in Fig. 3. Doing so gives the transmitter the necessary feedback to use rate adaptation algorithms and retransmit only the necessary frames. However, the overhead in terms of CPU capacity caused by BACK is significant as the number of BACKs and their requests increase linearly with the size of the multicast group. Thus, APs need more time to compute bitmaps and Cyclic Redundancy Codes (CRCs), which makes it impractical. For this reason, its implementation in market devices is very limited [9]. Thus, in this work, this policy is not considered.

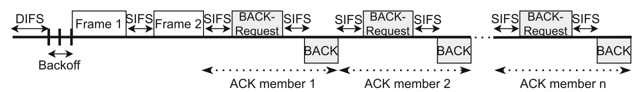


Fig. 3 Exchange of 2 multicast frames confirmed with Block-ACK. The frames are addressed to the multicast group's address, and then, upon request from the AP, each member sends a BACK confirming the reception of the frames

3.1.4 DMS

This extends the capabilities already introduced in the IEEE 802.11v amendment [10]. In a multicast group with n Multicast Receivers (MRs), DMS generates n copies of the multicast frame, assigning each copy as a unicast frame to individual Multicast Receivers (MRs), as illustrated in Fig. 4. Thus, multicast frames are transmitted in the same way as unicast frames. Like unicast transmissions, the multicast frames undergo retransmission until the source receives an acknowledgment (ACK) or the retransmission counter reaches its maximum limit. This approach provides multicast streams with a similar reliability as unicast streams. However, it consumes more airtime than Legacy. In addition, it is limited in scalability because the required resources increase linearly with the number of MRs. Unicast transmissions can use rate adaptation algorithms, so when multicast is converted into unicast, faster data rates can be used, frees up channel air time, making DMS very efficient in small groups.

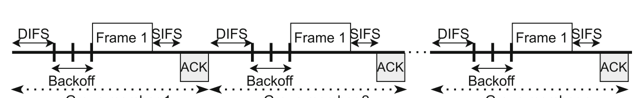


Fig. 4 Exchange of one multicast frame using DMS. The frame is replicated and sent as a unicast frame to each member. Each copy of the frame is addressed to each individual member

Table 1 Performance classification of the GATS

Policy	DMS	Legacy	GCR-UR
Scalability	Low	High	High
Delay	Variable	Moderate	Moderate
Efficiency	Variable	Low	Low
Reliability	High	Low	Moderate

In short, as shown in Table 1, DMS is most effective for small groups. It provides high reliability, but it suffers from poor scalability. Legacy displays low reliability but no scalability problems, while GCR-UR improves the reliability of Legacy at the cost of using more airtime, which impacts the unicast streams.

3.2 VHT Wi-Fi Networks

The IEEE 802.11ac amendment [11] has increased the maximum theoretical speeds up to 7 Gbps with the introduction of the VHT physical layer, increasing channel width to 80 MHz. Furthermore, IEEE 802.11ac also enables the use of 160 MHz channels, which can double the theoretical data rate. These wider channels are achieved through the combination of narrower bands, i.e., an 80 MHz channel is made up of two contiguous 40 MHz channels, which are formed by two contiguous 20 MHz channels each, where one of them is the primary, and the rest are secondary channels.

Another key feature of the new physical layer is the introduction of the 256-QAM modulation. This allows a big increase in the data rate with respect to the highest modulation in IEEE 802.11n [12], which was 64-QAM, reaching up to 867 Mbps per spatial stream when using the new 256-QAM, the compulsory maximum coding rate of 5/6 and short guard interval. In addition, IEEE 802.11ac introduces the support of up to 8 spatial streams, which results in a theoretical aggregate of 3.5 Gbps when using the 80 MHz channel width.

This dramatic increase in the data rate for multimedia transmissions has a positive effect on the Quality of Service (QoS) is positively affected. More in particular, this also alters how the aforementioned GATS should be used. While Legacy and UR keep transmitting at the basic rate, which keeps a high channel occupancy, DMS can now use much higher data rates as it uses rate adaptation algorithms. This partly mitigates its scalability problems. With this, it seems clear that the network conditions when using the different GATS are different in VHT networks compared with previous versions of the standard. For this reason, the performance evaluations existing in the literature have become obsolete. Thus, this is the first work that studies the performance of the GATS policies presented in the IEEE 802.11aa amendment on VHT Wi-Fi networks and provides recommendations for its use.

4 Evaluation of the IEEE 802.11aa amendment in VHT networks

Gathering large amounts of data from various scenarios becomes complicated on physical test beds, where the possible parallelization levels and the available time are limiting factors. For this reason, the evaluation of the standard carried out in this work uses the popular network simulator NS-3 [13], version 3.35. The original simulator does not implement the IEEE 802.11aa GATS policies, so this work has extended it to implement DMS, Legacy, and GCR-UR. The implementation of these policies is publicly available[1].

4.1 Simulated scenarios

The scenarios consist of a multicast group with a configurable number of MRs that receives a multicast stream through the video (VI) Access Category (AC). EDCA is used in all the simulations. A set of n unicast stations (STAs) introduces load in the network to simulate different channel conditions, with $n/2$ on the VI AC and the other $n/2$ on the Best Effort (BE) ACs. The deadline for all frames is 2 s, after which the frames are dropped. The multicast stream is generated by a server connected to an AP to which the MRs and the STAs are connected. The STAs and MRs are distributed in a random manner inside a circular area with a 30-meter radius, with the AP in the middle. In this way, the MRs and STAs have different Received Signal Strength Indicators (RSSIs). This server also receives the uplink traffic, as shown in Fig. 5. NS3-AI [14] uses shared memory to connect

[1] Available at: https://github.com/blasf1/ns-3-dev-git/tree/wifi_multicast_3.35

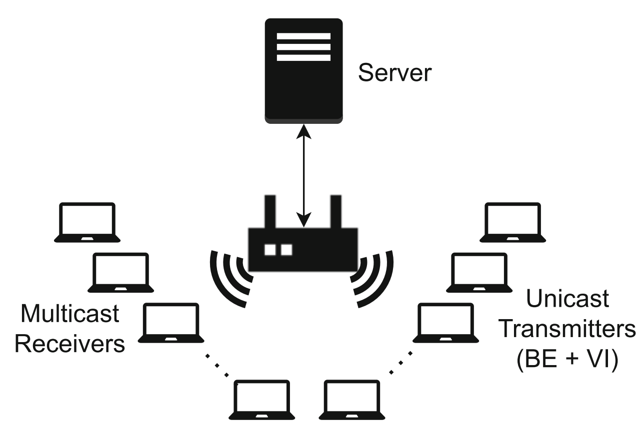


Fig. 5 Simulated topology

Table 2 Physical layer settings

Setting	Value
PHY layer	802.11ac
Rate adaption	Minstrel
Tx Power	15 dbm
Error rate model	Table-based
Max distance	30 m

NS-3 to external Python libraries. In this case, it is used to connect the simulation with an external script, where the Key Performance Indicators (KPIs) are gathered and processed. The AP sends a set of KPIs to the script every 0.25 s. The unicast data rate is set to 1 Mbps while the multicast data rate is set to 1.5 Mbps to match video call applications such as Microsoft Teams [15]. This is done to simulate scenarios such as hybrid lectures at a university or a student dorm. The simulations were carried out using the IEEE 802.11ac physical layer model of NS-3, and Minstrel was used for the rate adaptation of unicast and DMS transmissions. The channel width minimum is set to 40 MHz to facilitate the use of the new VHT MCSs. Contrary to previous evaluations, the scenario presented in this work uses Minstrel [1] as a rate adaption algorithm, representing more realistically the behavior of DMS and the channel occupancy. Since MRs are positioned randomly inside a 30-meter disc, the transmission power is adjusted to have stations with different levels of RSSI. This way, effects such as DMS stations using slow datarates are factored in. We use the table-based error rate model as it is the only one supported by the VHT physical layer in NS-3. The details on the configuration of the physical layer are summarized in Table 2.

When using simulations, there is a close relation between the results and the scenario. However, the number of different scenarios that can be tested is infinite. The scenarios designed in this section aim to overcome this by incorporating random positioning of the STAs into a realistic scenario. For this reason, we also base the selection of parameters such as the multicast transmission rate in real-life data.

A set of scenarios is defined to gain an understanding of the efficiency and the scalability of the GATS policies by varying the number of STAs (the load of the network) and the number of MRs (the size of the multicast group). Two blocks of scenarios have been defined: efficiency scenarios, summarized in Table 3, and scalability scenarios, summarized in Table 4. For all scenarios, we gather:

(i) the normalized multicast goodput, i.e., the amount of user information delivered correctly normalized with respect to the amount of information that was injected in the network;

(ii) the channel occupancy, i.e., the share of airtime being used for transmissions;

(iii) the share of retransmissions, i.e., the share of frames that are retransmitted over the total number of frames; and

(iv) the average delay experimented by the stations for the multicast transmissions.

The different scenarios evaluated are described in the following subsections.

4.1.1 Efficiency

An efficient GATS policy is essential in heavily loaded networks to avoid further channel congestion. A more efficient GATS policy does not necessarily reduce the occupancy, but it may be able to fit more frames into the available airtime, increasing throughput. Thus, to test the efficiency of the GATS policies, the number of MRs is locked, and two scenarios are defined: scenario 1, with a small multicast group ($m = 4$ MRs), and scenario 2, with a big multicast group ($m = 10$ MRs). This group size is determined by preliminary simulations, where groups bigger than $m = 10$ MRs do not show any further behavior changes. With these two multicast groups, the number of STAs is progressively increased to increment the load on the network. Thus, the number of STAs is given by $u : u = 4n : n \in [\![1, 9]\!]$, i.e., tests with all the number of STAs that are multiples of 4 between 4 and 36. The upper limit of the set of efficiency simulations has been determined empirically using test simulations to identify when frames start exceeding the deadline and are consequently dropped, indicating the maximum load level. Each simulation represented 30 s, but only the last 20 s were used for the evaluation. The first 10 s are disregarded as they are used as a stabilization period. Each of these combinations is repeated 10 times with different seeds.

4.1.2 Scalability

These tests aim to find the crossover point when the number of MRs makes DMS produce excessive overhead. The size of a multicast group influences the selection of the GATS policy. While Legacy and GCR-UR are independent of the group size, DMS's performance depends on it because the number of frames it transmits increases proportionally to the number of MRs in the multicast group. However, in small multicast groups, the use of DMS can be beneficial as the transformation of multicast transmissions into unicast transmissions makes it possible to use faster MCSs. Thus, to find this crossover point, three scenarios are defined where the network load (introduced by the STAs) is locked, and then for each of them, the number of MRs is gradually incremented. Scenario 1 has 8 STAs ($u = 8$) to simulate a low

Table 3 Evaluation parameters for the efficiency scenarios

Unicast STAs datarate	# Unicast STAs	# MRs	Multicast datarate
1 Mbps	$u : u = 4n; n \in [\![1, 9]\!]$	Scenario 1: $m = 4$ Scenario 2: $m = 10$	1.5 Mbps

Table 4 Evaluation parameters for the scalability scenarios

Unicast STAs datarate	# Unicast STAs	# MRs	Multicast datarate
1 Mbps	Scenario 1: $u = 8$ Scenario 2: $u = 20$ Scenario 3: $u = 28$	$2m : m \in [\![1, 8]\!]$	1.5 Mbps

load; scenario 2 has 20 STAs ($u = 20$) to simulate a medium load; and, finally, scenario 3 has 28 STAs ($u = 28$) to simulate a high load. These load levels are determined by looking at the results of the efficiency scenarios. Each of these scenarios tests the effect of different numbers of MRs. The set of tested group sizes is given by $2m : m \in [\![1, 8]\!]$, where m is the number of MRs.

4.2 Results Discussion

In this subsection, the results of the scenarios presented above are discussed, starting with the results of the efficiency scenarios first.

4.2.1 Efficiency

Fig. 6 shows the results of scenario 1 for the efficiency evaluation, particularly the scenario with a small multicast group. As expected, when the number of MRs is low enough, DMS achieves a higher multicast goodput regardless of the load in the network, as shown in Fig. 6a. Ninety-five percent confidence intervals for the 10 executions are shown as a shading that matches the corresponding line's color. The use of higher data rates enabled by the feedback available in DMS allows it to fit in more frames and make more efficient use of the airtime than Legacy and GCR-UR, which need to transmit at the basic rate. This causes each frame to take much more airtime than those transmitted with DMS. Moreover, DMS retransmits the frames that are lost thanks to the use of ACKs, thus improving reliability. Overall, GCR-UR achieves a higher multicast goodput than Legacy, thanks to the retries that deliver frames that Legacy could not deliver. However, there is a point where $x = 20$ at which Legacy performs better. This happens because Legacy delays the congestion in the network with respect to GCR-UR. In this way, the lower load of Legacy at this stage causes fewer collisions, and more frames are delivered even without retries. This can be seen by the share of retransmissions (share of received frames that were retransmission over the total number of received frames) in the unicast traffic shown in Fig. 6c. In the stages where the load is even higher, the network reaches its congestion limit, and the retries play a fundamental part again for GCR-UR, as those frames lost due to collisions have a second chance. Similarly, before $x = 20$, the low load in the network helps GCR-UR deliver more frames as most frames are lost due to noise in the channel, not collisions.

The capacity of DMS to fit in more frames of the multicast thanks to the higher data rates stream leads to a worsening of the unicast goodput (which is transmitted uplink). Similarly, the greater airtime consumed by GCR-UR also has a negative impact on the unicast traffic compared with Legacy. This is also reflected in the channel occupancy in Fig. 6b. A very important metric to take into consideration in video applications is delay. The delay provides a good understanding of the behavior of the different EDCA queues. Figure 6d shows how DMS frames spend, on average, much longer in the queue as the load increases. This happens because DMS injects m times more frames than Legacy and $m/2$ times more frames than GCR-UR. Consequently, more frames need to wait to gain medium access, and the queue gets fuller. However, note that DMS does not reach the deadline with the group size tested in scenario 1. This is expected as more frames in the downlink compete to use the channel, affecting the uplink. Something similar happens in the delays with both GCR-UR and Legacy. The fact that Legacy injects half the number of frames gives it an advantage in terms of delay, although this does not result in a higher goodput as it offers less reliability.

When the bigger multicast group is used in scenario 2, the results are notably different, as shown in Fig. 7, especially for DMS. The presence of a larger number of MRs multiplies the number of frames that DMS needs to send, as it transmits a unicast frame per MR. This causes congestion in the channel, and its performance degrades quickly as the load in the network increases, as shown in Fig. 7a, where the multicast goodput falls behind GCR-UR and Legacy after only 12 unicast stations. Even if frames transmitted using DMS use higher data rates and less airtime, the fact that there are more MRs in this scenario causes a bigger overhead. Sending m faster frames plus their respective contention periods uses more airtime than sending a single slower frame with Legacy or two slower frames in GCR-UR. The greater congestion caused by DMS obviously has a negative impact on the num-

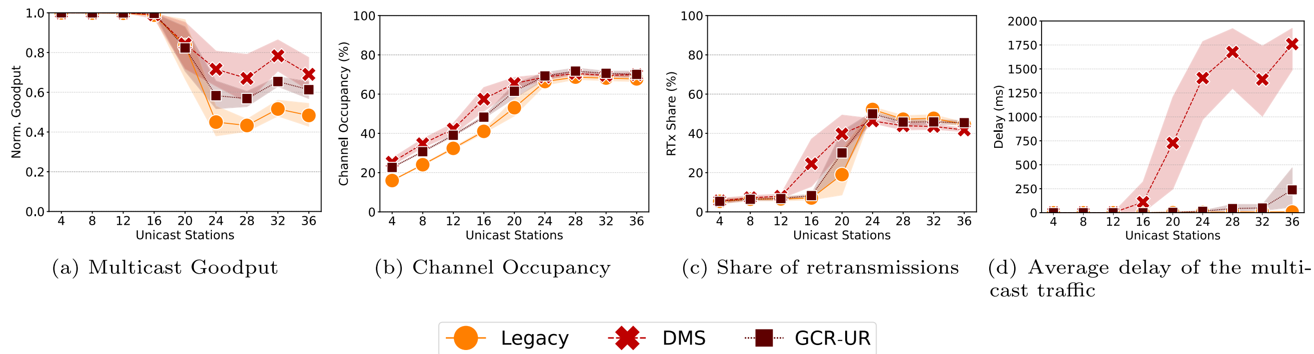


Fig. 6 Scenario 1 of the efficiency evaluation ($m = 4$ MRs) with unicast sources of 1 Mbps

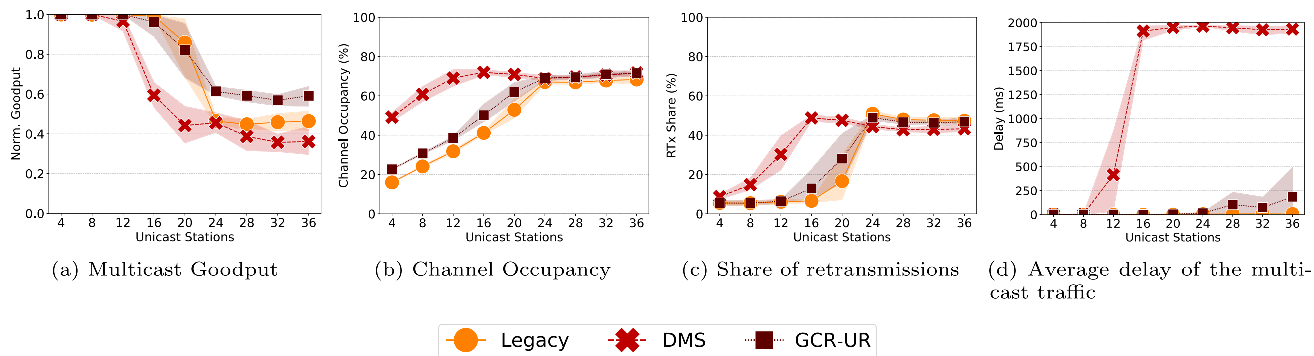


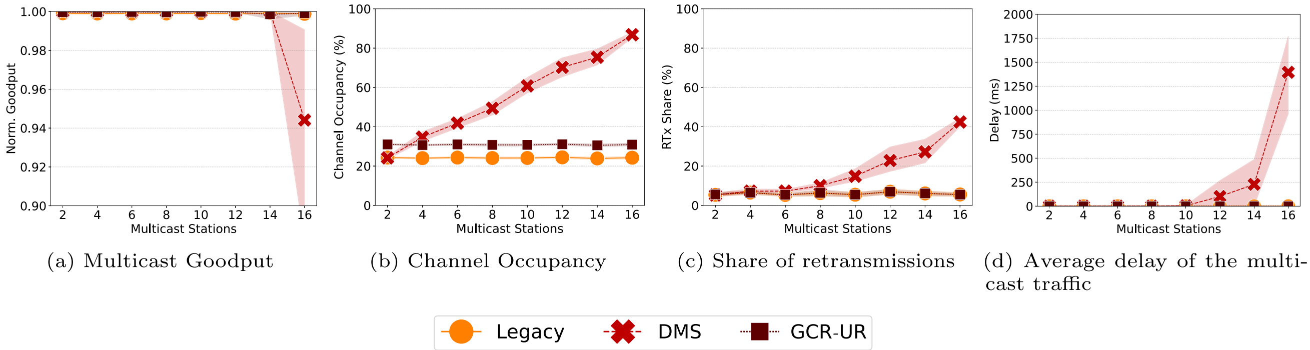
Fig. 7 Scenario 2 of the efficiency evaluation ($m = 10$ MRs) with unicast sources of 1 Mbps

ber of collisions, as shown in Figure 7c. Note that even if the share of retransmissions includes only unicast traffic, it gives a good enough idea of when saturation on the network is reached and the number of collisions happening in the channel. The behavior of Legacy and GCR-UR is virtually the same as in scenario 1, as the group size does not affect these policies. Channel occupancy, shown in Fig. 7b, perfectly reflects the overload of the channel caused by DMS. Multicast delay is also negatively impacted in DMS, as this time, contrary to what happened in scenario 1, frames are exceeding the deadline and being dropped, as shown in Fig. 7d.

4.2.2 Scalability

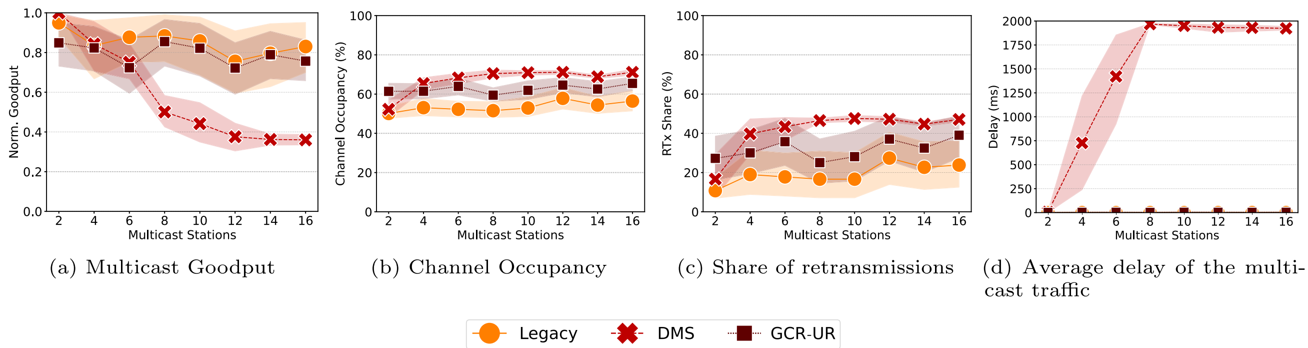
Regarding scalability, Legacy and GCR-UR are expected to show stable behaviors as they are not affected by the number of MRs. However, DMS's performance worsens as the number of MRs increases, as shown in Fig. 8. Legacy and GCR-UR remain stable, delivering almost every frame. On the other side, DMS's goodput falls with bigger multicast groups, as shown in Fig. 8a. However, before reaching the point where performance drops, DMS can deliver 100% of the traffic thanks to the existence of feedback. In contrast, GCR-UR and especially Legacy fail to do so, even though they are still very close to 100%. Figure 8b shows how the channel occupancy increases progressively with the number of MRs when DMS is used. Consequently, as the channel becomes more congested, the goodput falls due to the increase in the number of retransmissions, as shown in Fig. 8c. The higher occupancy of the channel causes the frames to wait longer in the queue until gaining medium access, and therefore, the multicast delay also increases, as shown in Fig. 8d.

Figure 9 shows the scalability scenario 2, i.e., with a medium load. This time, DMS's multicast goodput degrades much faster and is already below Legacy for $x = 6$ MRs, as shown in Fig. 9a. The higher load in the network negatively influences the number of MRs that DMS can handle. In this scenario, the network load brings it closer to saturation. Thus, when DMS needs to send frames to a reasonable number of MRs, the higher number of frames trying to gain channel access ends up saturating the channel and making the performance drop. Legacy achieves a higher multicast goodput than GCR-UR, despite the latter retransmitting all frames, which should increase the chances of frames being delivered. However, the network in this stage is getting close to saturation, and the extra airtime used by GCR-UR retransmissions is causing a higher occupancy, as shown in Fig. 9b. DMS's occupation increases until it reaches saturation. On the other hand, GCR-UR and Legacy, while close to saturation, stay



(a) Multicast Goodput (b) Channel Occupancy (c) Share of retransmissions (d) Average delay of the multicast traffic

Fig. 8 Scenario 1 of the scalability evaluation ($u = 8$ STAs) with unicast sources of 1 Mbps



(a) Multicast Goodput (b) Channel Occupancy (c) Share of retransmissions (d) Average delay of the multicast traffic

Fig. 9 Scenario 2 of the scalability evaluation ($u = 20$ STAs) with unicast sources of 1 Mbps

below, which allows them to reduce the number of collisions, as shown in Fig. 9c, which allows the increase in the multicast goodput mentioned above. It also shows how the extra airtime used by the retransmissions of GCR-UR increases the probability of collisions, which explains its lower goodput, as the higher chances of delivery due to the existence of retransmissions cannot compensate for the higher chances of collisions, caused by the close-to-saturation scenario caused by such retransmissions. The overhead caused by DMS has a big impact on the delay of multicast traffic as well, as shown in Fig. 9d. Legacy and GCR-UR do not reach saturation, and the frames can gain medium access without accumulating in the queue, which keeps delays low. DMS generates too many frames for them to fit within the available airtime. Thus, frames pile up in the queue and end up being dropped as they exceed the deadline.

Finally, Fig. 10 shows scenario 3, representing a high-load situation. The saturation is clearly shown by the occupancy and the share of retransmissions in Figs. 10b and 10c. When the number of MRs is smaller than 6, DMS benefits from faster data rates and feedback, which allows it to fit in more frames and deliver them. However, as the network is saturated, the performance of all the policies is negatively affected, with a noticeable decrease in the multicast goodput with respect to scenario 2, as shown in Fig. 10a. Legacy performs worse than GCR-UR again in this case, as saturation occurs with both policies. Then, the retries of GCR-UR increase the probability of successful delivery. In DMS, many more frames are generated, and thus, they have to wait longer in the queue to gain medium access. Due to the high level of congestion, many of them are dropped as they cannot gain medium access before the 2-second deadline, as shown in Fig. 10d.

4.3 Usage Recommendations

Some general usage recommendations can be drawn from the results obtained in the evaluation. Legacy generally offers poor reliability, but when the channel occupancy is close to saturation and with bigger multicast groups, it can avoid further channel saturation compared with GCR-UR and DMS. Avoiding saturation is essential for maintaining performance, making it perfect for this situation. Thus, Legacy is most useful in networks with significant loads that do not often reach saturation. In HT networks, transmitting at a basic rate was more detrimental to Legacy as the rest of the stations were also transmitting at lower data rates, which made it more difficult for them to fit their traffic in the channel.

DMS offers the best reliability and efficiency with smaller multicast groups. The exact number of MRs depends on the load of the network and the number of STAs expected to use the network since more STAs saturate the network more,

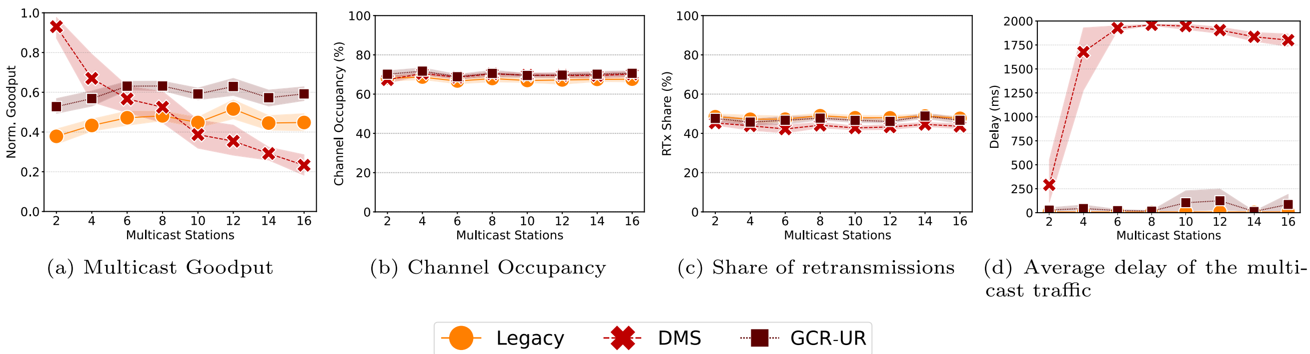


Fig. 10 Scenario 3 of the scalability evaluation ($u = 28$ STAs) with unicast sources of 1 Mbps

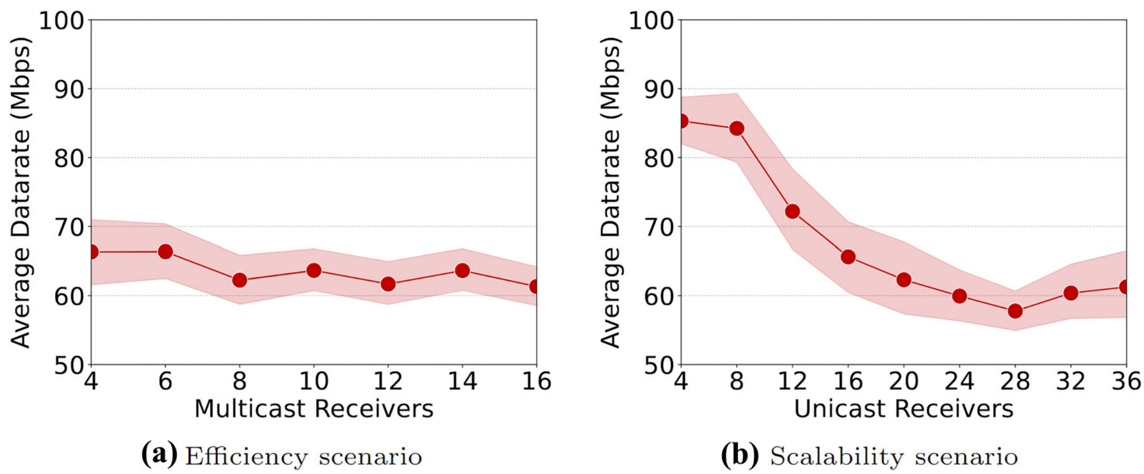


Fig. 11 Average datarate of the multicast stations when using DMS

even with the same total amount of injected traffic, but in general, DMS does not cope well with groups of more than 4 MRs. This number may be higher with lower loads. DMS proves more versatile in VHT networks as the higher available data rates enable it to fit in the frames of more MRs in the same airtime. Allowing DMS to use rate adaptation algorithms gives it a further advantage. Thus, it is possible to use DMS in VHT networks where the link quality is usually good, and therefore, higher data rates can be used, and multicast groups are predicted to be small. The average data rate used by the MRs when using DMS is shown in Fig. 11. When the network is saturated, GCR-UR performs better as the retries increase reliability. Frames have more chances of being delivered despite the high levels of collisions as it is not that frequent that a frame and its retry get lost due to collisions. Thus, GCR-UR is more suitable for networks that often operate under saturation.

Another aspect to consider when selecting the appropriate policy is the capacity of the AP. Not all policies require the same computing resources. DMS demands more resources, as it involves the AP replicating the same frame with different headers as many times as MRs there are in the group, having to process their respective ACKs as well. On the other hand, Legacy does not generate any extra overhead as it sends a single frame to all MRs, regardless of the number of MRs. GCR-UR is the middle ground, as it has to replicate the data frames as many times as retries, regardless of the number of members in the multicast group. Moreover, the headers do not need to be modified, and there is no need to process the ACKs.

A method to dynamically select the transmission policy to the channel conditions and the number of MRs would be ideal for extracting the maximum performance of the network at each moment in time. We build on this work to present such a mechanism in [16].

5 Conclusion

This paper presents the first performance evaluation of the GATS policies introduced in the IEEE 802.11aa amendment on a VHT physical layer that uses a rate adaptation algorithm to recreate realistic conditions. Our evaluation shows considerable differences in the performance of the GATS policies due to the higher available bandwidth with respect to previous HT networks. The literature shows how the overhead introduced by DMS with the lower data rates provided by HT networks renders it hardly usable. However, in VHT networks, it is the most efficient policy with smaller groups and even with middle-size groups if the network's load is low. Our evaluation shows that channel occupancy is the main KPI affecting the performance of all the policies, with the number of MRs also being relevant in DMS. A dynamic approach that chooses the best GATS policy for each situation would be able to achieve a sustained performance by combining the advantages of all of them.

Author Contributions All authors participated in the conceptualization. BG wrote the original draft of the manuscript and designed the software and data visualization. BG, EC, and JV designed the methodology and the analysis. All authors reviewed and edited the final manuscript. EC, JV, and AG supervised and acquired funds.

Funding This work has been supported by MCIN/AEI/10.13039/501100011033 (ERDF "a way of making Europe") under grants PID2022-142332OA-I00 and PID2021-123627OB-C52. This work is also funded by UCLM and the European Social Fund (Grant 2019-PREDUCLM-10921), the Government of Castilla-La Mancha (project SBPLY/21/180501/000195), and UCLM (Project 2023-GRIN-34056). This work is also supported by the EU "NextGenerationEU/PRTR", MCIN, and AEI (Spain) under Project IJC2020-043058-I. The authors have not disclosed any funding.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Xia, D., Hart, J., & Fu, Q. (2013). Evaluation of the Minstrel rate adaptation algorithm in IEEE 802.11g WLANs.In: *Proceedings of IEEE ICC*, Budapest, Hungary.
2. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) (2012). Specifications Amendment 2: MAC Enhancements for Robust Audio Video Streaming. *IEEE Std 802.11aa-2012*, pp. 1–162.
3. Banchs, A., de la Oliva, A., Eznarriaga, L., Kowalski, D. R., & Serrano, P. (2014). Performance analysis and algorithm selection for reliable multicast in IEEE 802.11aa wireless LAN. *IEEE Transactions on Vehicular Technology, 63*(8), 3875–3891.
4. de la Oliva, A., Serrano, P., Salvador, P., & Banchs, A. (2013). Performance evaluation of the IEEE 802.11aa multicast mechanisms for video streaming. In: *Proceedings of IEEE WoWMoM*, Madrid, Spain.
5. Santos, M. Á., Villalón, J., & Orozco-Barbosa, L. (2012). Evaluation of the IEEE 802.11aa group addressed service for robust audio-video streaming. In: *Proc. of IEEE ICC*, Ottawa, Canada.
6. Salvador, P., Cominardi, L., Gringoli, F., & Serrano, P. (2013). A first implementation and evaluation of the IEEE 802.11aa group addressed transmission service. *Computer Communications Review, 44*(1), 35–41.
7. Gringoli, F., Serrano, P., Ucar, I., Facchi, N., & Azcorra, A. (2019). Experimental QoE evaluation of multicast video delivery over IEEE 802.11aa WLANs. *IEEE Transactions on Mobile Computing, 18*(11), 2549–2561.
8. Mansour, K., Jabri, I., Ezzedine, T. (2018). IEEE 802.11 aa GATS mechanisms performance study for very high throughput WLANs. In: *Proceedings of IEEE WCNC*, Barcelona, Spain.
9. Wi-Fi network roaming with 802.11k, 802.11r, and 802.11v on iOS, iPadOS, and macOS. Accessed: 2023-05-13. [Online]. Available: https://support.apple.com/en-bh/HT202628
10. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) (2011). Specifications amendment 8: IEEE 802.11 wireless network management. *IEEE Std 802.11v-2011*, pp. 1–433.
11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) (2013). Specifications amendment 4: enhancements for very high throughput for operation in bands below 6 GHz. *IEEE Std 802.11ac-2013*, pp. 1–425.
12. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) (2009). Specifications Amendment 5: Enhancements for Higher Throughput. *IEEE Std 802.11n-2009*, pp. 1–565.
13. Riley, G. F., Henderson, T. R. (2010). The NS-3 Network Simulator. In: Modeling and Tools for Network Simulation, pp. 15–34.
14. Yin, H., Liu, P., Liu, K., Cao, L., Zhang, L., Gao, Y., & Hei, X. (2020). Ns3-Ai: fostering artificial intelligence algorithms for networking research. In: *Proceedings of Workshop on NS3*, Gaithersburg, MD, USA.
15. Microsoft (2023). Microsoft Teams—Network requirements. Accessed: 2023-05-13. [Online]. Available: https://docs.microsoft.com/en-us/microsoftteams/prepare-network
16. Gómez, B., Coronado, E., Villalón, J., & Garrido, A. (2022). Intelli-GATS: Dynamic selection of the Wi-Fi multicast transmission policy using interpretable-AI. *Wireless Communications and Mobile Computing, 2022*, 7922273.

Blas Gómez is a PhD student at Universidad de Castilla-La Mancha since 2020. In 2019, he completed his MSc in Computer and Networks Engineering from the Polytechnical University of Valencia and his BSc in Computer Science in 2018 from Universidad de Castilla-La Mancha. His current research interests include the delivery of multimedia content over SDWLANs, wireless communications, MEC systems, and AI-driven network management.

Estefanía Coronado is a Senior Researcher at Fundació i2CAT (Spain) and a Juan de la Cierva Senior Researcher at the University of Castilla-La Mancha (Spain). From 2018 to 2020 she was an Expert Researcher at Fondazione Bruno Kessler (Italy). In 2018, she completed her PhD at the University of Castilla-La Mancha (Spain) on multimedia ML-driven distribution over SD-WLANs. She has worked in several H2020 projects, acted as technical task leader, and as PI of a national project. She has published around 50 papers in international journals and conferences. Her current research interests include wireless communications, MEC systems, and AI-driven network management.

José Villalón is currently an Associate Ph.D. Professor at UCLM. He received an M.Sc. degree in Computer Science and a Ph.D. in Computer Engineering from the University of Castilla-La Mancha, Spain, in 2003 and 2007, respectively. He has been a Visiting Researcher at INRIA, France. His research interests include high-performance networks, wireless networks, QoS and QoE over IEEE 802.11 and WiMAX, multicast transmission, software-defined networking, video distribution, edge computing and error-resilient protocol architectures.

Antonio Garrido received the Ph.D. degree from the University of Valencia, Spain, in 1991. In 1986, he joined the Computer Systema Department at the University of Castilla-La Mancha, where he is currently a Full Professor of Computer Architecture and Technology. His research interests include wireless sensor networks, video compression, and video transmission. From the year 2000, collaborates with the National Agency for Quality Assessment and Accreditation of Spain and several regional quality agencies in Spain.