

This is a postprint version of the following published document:

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.

DOI: [10.1109/tmc.2018.2876000](https://doi.org/10.1109/tmc.2018.2876000)

© 2019, IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Experimental QoE evaluation of multicast video delivery over IEEE 802.11aa WLANs

Francesco Gringoli, *Senior Member, IEEE*, Pablo Serrano, *Senior Member, IEEE*, Iñaki Ucar, *Student Member, IEEE*, Nicolò Facchi, *Member, IEEE* and Arturo Azcorra, *Senior Member, IEEE*

Abstract—The IEEE 802.11aa amendment standardised the Group Addressed Transmission Service (GATS), which extends 802.11 WLANs with a novel set of MAC mechanisms to support an effective and efficient multicast video service. The key challenge with GATS is the selection of the best scheme and its configuration for a given network scenario, as the standard does not provide any guidelines nor any assessment of the performance of each mechanism. Although some previous studies have addressed this challenge, their evaluation is either via analysis or simulations under non-realistic assumptions, or based on Quality of Service (QoS) metrics instead of video quality metrics, which are required for a proper video performance assessment. In this paper, we deploy a mid-size real-life testbed and develop a thoughtful methodology to perform an extensive Quality of Experience (QoE) evaluation of GATS under a variety of scenarios. We analyse the performance of the novel schemes under ideal conditions, as well as under controlled and non-controlled interference, assessing their ability to provide an adequate QoE and quantifying the resources left for other type of traffic. Ours is the first thorough QoE evaluation of GATS in a real-life scenario, providing key insights on their performance, and can be used to derive configuration guidelines for the schemes.

Index Terms—WLAN, 802.11aa, GATS, QoE, QoS, Multicast, Video streaming, VMAF

1 INTRODUCTION

The Group Addressed Transmission Service (GATS) of the IEEE 802.11aa amendment [1] supports a more effective and efficient delivery of multicast streams over WLANs. Effectiveness is improved thanks to the use of retransmissions (either with “preemptive” retransmissions or using ARQ), while efficiency is improved thanks to the use of higher Modulation and Coding Schemes¹ (MCS).² One of the main motivations behind GATS is the growth of real-time multimedia traffic, in scenarios such as e.g. sports events (i.e., stadiums), set-top boxes, teaching environments or enhanced-driving applications (cameras at intersections, see-through applications).

In fact, video delivery over wireless networks (both unicast and multicast) has received a notable attention from the research community (we review the state of the art in Section 2), due to the involved research challenges: the lossy medium, the non-straightforward relation between these

losses and the quality of experience (QoE), or the best approach to recover from these losses, including schemes such as, e.g., leader-based protocols, which are not supported by the standard and, therefore, are hardly likely to be widely deployed. In contrast, GATS provides a set of standardised mechanisms that build on existing 802.11 procedures and whose practicality is hence less questionable.

The diversity of mechanisms introduced by GATS (we summarise them in Section 2) makes the selection of the best performing configuration given a network setting a tough task: the standard, in fact, does not give any insight into the performance of each mechanism, nor provides any guideline on which mechanism to use for a given scenario. Indeed, in our previous analytical [3] and experimental [4] studies, we have confirmed that each GATS mechanism provides a different performance trade-off in terms of complexity, efficiency and effectiveness, and therefore the “optimal” scheme depends on the parameters of the scenario. One key limitation of these studies, though, is that the performance evaluation is based on metrics such as saturation throughput or packet losses, and therefore its mapping to QoE (which is particularly relevant in the case of video) is not clear.

In this paper, we perform an extensive evaluation of the performance of GATS when delivering multicast traffic in a real-life scenario. To this aim, we deploy a mid-size testbed consisting of 25 WLAN stations implementing GATS, and perform a thorough evaluation of video delivery using a carefully-crafted methodology under a variety of scenarios, including a standardised set of videos covering different types of visual content, the cases of controlled UDP and TCP interference, the use of EDCA differentiation, and performance under uncontrolled (i.e., “real”) interference during a typical working day (morning-night). Given the tremendous

- F. Gringoli is with the Department of Information Engineering, CNIT/Univ. of Brescia, Brescia, Italy.
E-mail: francesco.gringoli@unibs.it
- N. Facchi is with the Department of Information Engineering and Computer Science, University of Trento
E-mail: nicolo.facchi@unitn.it
- P. Serrano, I. Ucar and A. Azcorra are with the Department of Telematic Engineering, Univ. Carlos III de Madrid, 28911 Leganés, Spain.
E-mail: {pablo, azcorra}@it.uc3m.es, inaki.ucar@uc3m.es
- A. Azcorra is also with Institute IMDEA Networks, 28912 Leganés, Spain.

1. Even though the term MCS was introduced to define the modulations available with HT- and VHT-PHY, throughout the paper we will use this term also to refer to “data-rates”, as defined in the OFDM-PHY document.

2. In the legacy multicast service, transmissions are performed with one of the ‘Basic Service Set’ rates, which decreases the overall performance of the WLAN because of the *performance anomaly* [2].

amount of experiments considered, a subjective evaluation is impractical, and therefore we rely on the reconstruction of trace-files and a recent tool developed by Netflix to estimate the video quality.³ More specifically, our key contributions are as follows:

- We describe the deployment of a real-life scenario for QoE-based evaluation, including the required software modules and tools for the generation, streaming and QoE evaluation.
- We design a thorough methodology for QoE evaluation, building on a standardised set of heterogeneous videos and the off-line processing of the trace-files collected during the experiments.
- We extensively assess the performance of GATS in a real-life scenario under no interference and under increasing UDP interference, evaluating the QoE and efficiency of each mechanism. We perform the same analysis with TCP traffic.
- We evaluate how the usage of 802.11e EDCA differentiation impacts the general performance when there is interfering data traffic, showing a very little improvement of the video quality at the cost of a marginal reduction of resources for best effort data traffic.
- Finally, we evaluate the performance of GATS over an entire working day (i.e., from 8am to 12pm, when neighbouring traffic is very significant) under uncontrolled (i.e., real) interference over four different 802.11 channels.

Our results provide valuable insights on the performance of GATS in a variety of scenarios, and can therefore be used to derive optimal configuration guidelines for 802.11aa networks. Furthermore, the methodology is easily repeatable, as all software used, source code and video traces are freely available, and therefore other researchers and practitioners can build on our results to derive experimental-driven configuration guidelines tailored to their scenario of interest.⁴

The rest of this paper is organised as follows. In Section 2 we summarise the operation of GATS and related work on this area. In Section 3 we describe our testbed and the methodology we follow to compute the QoE when streaming multicast video. In Section 4 we analyse the performance of GATS under controlled interference, this including no interference, increasing UDP interference, TCP traffic and the impact of EDCA differentiation. In Section 5 we report the performance in the wild during a working day, running the experiments under uncontrolled interference generated by external users. Finally, Section 6 concludes the paper.

2 BACKGROUND AND RELATED WORK

2.1 The IEEE 802.11aa GATS in a nutshell

GATS introduces a set of MAC schemes to improve the effectiveness and efficiency of multicast transmission over

3. In this way, ours is not a subjective evaluation. However, given that this tool provides state-of-the-art correlation with human perception, for the sake of readability throughout the paper we will refer to our evaluation as “QoE-based” evaluation.

4. The traces from our experiments and scripts used to generate them are available at <http://netweb.ing.unibs.it/openfwf/GATS>

WLANs, namely, Directed Multicast Service (DMS) and GroupCast with Retries (GCR). To summarise them, assume a WLAN where a multicast flow is sent to a set of receivers. With the **legacy** scheme, illustrated in Fig. 1 (top), the procedure is: for each multicast packet, an independent backoff procedure is triggered, and only one transmission attempt is made. Because of this, the card chooses a relatively low (i.e., robust) MCS, to maximise the chances that it is successfully received by all destinations, and therefore each transmission takes a relatively long time, as illustrated in the figure.

The **DMS**, not illustrated in Fig. 1, was firstly introduced in 802.11v and consists of performing a standard unicast transmission for each packet and receivers. Although this mechanism maximises the delivery probability, it poses severe scalability issues as we showed in our previous work [3], [4] and therefore we will not further consider it in this paper.

The GCR offers two alternatives, Unsolicited Retries (UR) and Block Acknowledgement (BA). The **UR** scheme (middle of Fig. 1) is a two-fold extension of the legacy mechanism, building on unacknowledged transmissions: on the one hand, the transmission can be done with any MCS (not necessarily from the “Basic Service Set”, i.e., the lower ones); on the other hand, each packet is transmitted (with a backoff procedure) a configurable number of times R , thus increasing the delivery ratio.

The **BA** scheme (Fig. 1, bottom) extends the Block Ack operation of the standard to support multiple destinations. Basically, its operation consists on the transmission of a “burst” of packets after the initial backoff procedure, and then a per-receiver polling operation to identify which packets have to be retransmitted. With this scheme, there are basically two main degrees of freedom: the MCS for the data packets and the initial burst length, which for simplicity we will refer hereafter as of length N packets.⁵ Differently than the other scheme, BA adds control overhead for collecting feedbacks which increases with the number of clients: using, e.g., theoretical models such as [3], the maximum number of stations supported with the 6 Mbps MCS under ideal conditions (i.e., no loss, video-only traffic) is approx. 200, which might preclude its usage in very-dense scenarios.

2.2 Related work

We next review the most relevant studies related to our work. Firstly, we consider those that address the general challenge of **multicast in 802.11 WLANs**, which has received a notable attention from the research community⁶ (see [5] for a general survey, and [6] for a survey with a focus on 802.11aa and its use with 802.1q). Here, one of the most related work is our previous report on the implementation on GATS [4], which includes an experimental evaluation that extends our initial simulation-based evaluation [7], but

5. To be more precise: (i) the MCS for the control packets of the polling procedure could be another degree of freedom, but for simplicity we fixed it to 6 Mbps; (ii) in the standard, the AP can transmit for as long as the TXOP parameter indicates. For practical considerations, our implementation translates this parameter into the length N .

6. We note that it is also attracting attention from the IETF, which has recently started exploring the performance issues of multicast over IEEE 802.11 wireless networks <https://www.ietf.org/archive/id/draft-perkins-intarea-multicast-ieee802-01.txt>

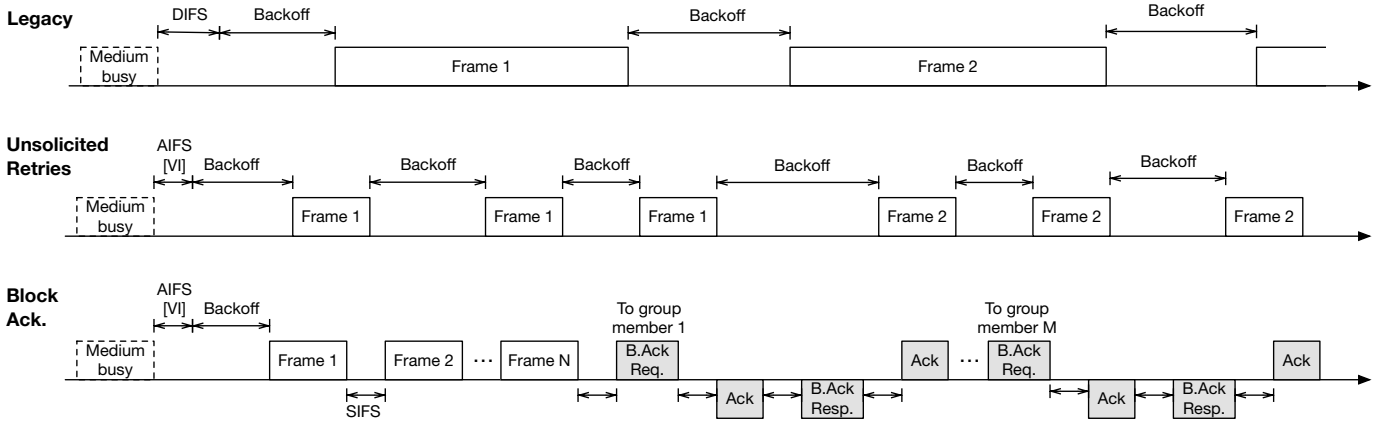


Fig. 1: Considered multicast mechanisms.

it mainly consists on traffic generators (instead of real video) and Quality of Service (QoS) metrics such as throughput or delivery ratio, but do not consider the impact of the configured MCS, of the EDCA differentiation, or the presence of TCP traffic as we do in this paper. Another simulation study is [8], which compares the performance in terms of PSNR and airtime of GATS UR and BA, as well as other schemes. Other relevant studies, based on theoretical analyses, are [9], [10] (for the case of the GCR-BA scheme), our previous work [3] (for all schemes), and the recent study [11] (for the case of 802.11ac PHY layer). Although these studies support the derivation of guidelines for the configuration of the multicast scheme, they assume constantly-backlogged multicast traffic (i.e., saturation), which is hardly the case when the aim is to guarantee video quality (as we will confirm with our experiments). It is worth noting that all these studies show that, in general, GATS improves multicast delivery, with the different schemes achieving different trade-offs between reliability, efficiency and complexity.

We next review those studies proposing **novel schemes for multicast**, which are typically based on collecting feedback from the receivers using an *ad hoc* protocol, that is programmed at the application layer [12], [13] or at the MAC layer [14]. Some other proposals advocate for the integration of Automatic Repeat Request (ARQ) and Forward Error Correction (FEC) mechanisms [13], [15]. Finally, it is worth mentioning two extensions for GATS: a leader-based variation of GCR-BA, where BA requests are sent only to a subset of receivers [16], and the use of packet aggregation schemes (A-MSDU and A-MPDU) for GATS operation [17]. The main drawbacks of these proposals are that in most cases their performance has not been assessed with real-life prototypes, and that they are unlikely to be broadly adopted by vendors as they are based on non-standard behaviour.

Other approaches have addressed the challenge of **adapting the video transmission** to the estimated conditions. A large amount of early work explored this topic mainly focusing on scalable video coding and traffic differentiation [18]–[22], with the main idea being to split users in different groups and, based on their channel quality, assign different parameters (e.g., MCSs and FEC rate) to different layers of scalable video. However, these studies are based on simulations and therefore suffer from the same practical

considerations considered above. The use of the EDCA differentiation is considered in [19], but the performance evaluation does not take into account any QoE metric. Additionally, a leader-based application-layer mechanism to adapt the frame rate is proposed in [23], but its performance evaluation is based on only three receivers and does not consider QoS.

Finally, it is worth mentioning a few papers that do not specifically address multicast but consider the case of **perceived video quality**. In [24] and [25], authors analyse the impact of different MAC parameters on MOS (Mean Opinion Score) performance. Similarly, [26] and [27] also address QoE evaluation of video: the former explores techniques for inferring the QoE of YouTube encrypted videos building on metrics such as delay and the frequency and duration of playback stalls, while the latter evaluates the SSIM (Structural Similarity) index [28].

To sum up, all previous studies suffer from at least one of the following limitations: 1) they do not consider standard schemes, but ad-hoc extensions that are unlikely to be adopted by hardware vendors; 2) their performance evaluation is based on simulations or analytical modelling, which neglects the impact of real-life conditions on performance; and/or 3) the evaluation is based on QoS metrics, lacking a proper assessment of the resulting quality of experience. In contrast to these limitations, ours is an extensive evaluation of the mechanisms included in the IEEE 802.11aa amendment, building on a real-life test-bed consisting of 25 commercial, off-the-shelf devices, which is based on a sound methodology to assess the quality of experience resulting from the streaming of a standardised set of videos.

3 TESTBED AND METHODOLOGY

In this section we describe the testbed we use throughout the paper to run the experiments, including the software employed, and then we detail the methodology that we follow to perform the QoE-based evaluation.

3.1 Physical deployment

Our testbed is located in the Dept. of Information Engineering of the University of Brescia, and is depicted in Fig. 2.

The testbed coexists with the corporate WiFi network and its dynamic user base of faculty members and students that generate heterogeneous interferences, with significant variations throughout the day as we will confirm with our experiments. In particular, the testbed consists of a set of devices for video transmission and reception, a set of devices for the generation of interference, and one device for controlling the experiments and processing the videos.

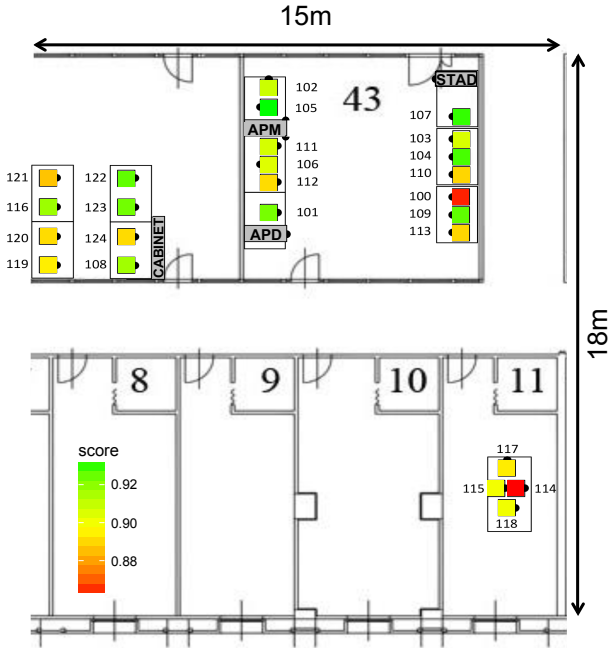


Fig. 2: Experimental setup. TCP downlink quality.

The devices involved in the video traffic are: (i) One desktop machine (single-core Pentium 4 CPU 3GHz, 1 GB RAM) that acts as AP, located in the top part of Fig. 2 (labeled as APM, i.e., AP for Multicast). It is equipped with a Broadcom BCM94318MPG 802.11b/g chipset and two 2 dBi omnidirectional antennas. (ii) 25 Alix 2d2 devices (single core Geode LX800 AMD 500 MHz CPU, 256 MB RAM), acting as a GATS-enabled wireless clients, which are spread through the deployment as illustrated in Fig. 2 (labeled with numbers from 100 to 124). Each device is equipped with the same wireless chipset as the APM and a 2 dBi omnidirectional antenna. Although such clients are relative low-powered, we confirmed throughout our experiments that they never exceeded their computational limits.

The generation of controlled interference relies on two additional desktop machines with similar characteristics to the APM node. One machine acts as AP and is placed slightly below the APM node in Fig. 2 (we mark it as APD, AP for Data), and the other machine acts as station and is located in the top-right of the figure (marked as STAD, STATION for Data). Both nodes embed an Atheros AR928X card, 802.11bg compatible. Finally, the control of the experiments and trace processing is done with a server (16-core Xeon E5-2630-v3 3.2GHz, 128 GB RAM), which is not showed in the figure.

All wireless nodes use the same transmission power, 17 dBm. Given the node placement, this causes dissimilar link qualities between each station and the Access Point. To

quantify this link heterogeneity, we run a set of experiments during night-time where we activate the `minstrel` Rate Adaptation (RA) algorithm, and transmit 60 s unicast TCP data from the APM node to each one of the 25 video-receivers in a round-robin manner across stations and channels $\{1, 6, 11, 14\}$. We then collect for each experiment the RA statistics (for each MCS_i , the total number of successful packets S_i and transmitted packets T_i), and we compute the link score as an average success ratio (total number of successes over total number of attempts) weighted by the MCS data rate value and normalised to the maximum MCS (i.e., 54Mbps):

$$\text{score} = \frac{\sum_i S_i \cdot MCS_i}{\max\{MCS\} \cdot \sum_i T_i} \quad (1)$$

According to this definition, a perfect channel with 100% success rate at the maximum MCS gets '1', while the worst channel obtains '0'.

We represent the obtained values averaged over the four channels in the heatmap illustrated in Fig. 2. As expected, despite the received quality is relatively good (all values are well above 0.8), there is a notable heterogeneity across links, which will affect the received video quality at each station.

Finally, we want to underline that with our testbed we investigate only 11g encodings for two main reasons: i) the coexisting university network works only in the 2.4GHz band and all its Access Points are 11g only; and ii) we focus our analysis on MAC timing issues and we leave further investigation involving PHY mechanisms as a future work (i.e., HT-PHY encodings, impact of spatial-stream and channel bonding). We are positive, though, that our results in controlled scenarios would have been relatively similar had we run the experiments in the 5GHz using 11a encodings, given that the PHY and MAC are practically the same. More specifically, the lack of backwards-compatibility mechanisms [29], [30] and lower interference from neighbouring traffic would have lead to better performance, despite the higher attenuation that could be addressed with a careful AP deployment planning.

3.2 Software used and streaming methodology

We next detail the software used and the complete methodology we designed to evaluate the performance of a given scheme to transmit video to a set of receivers, which is illustrated in Fig. 3.

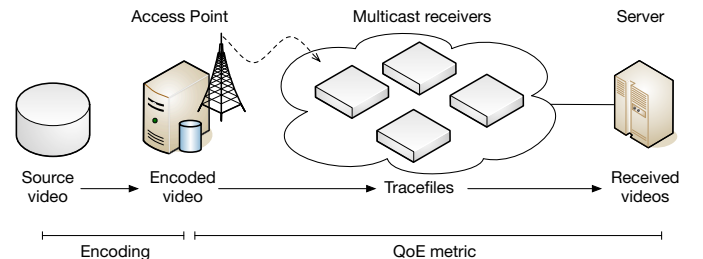


Fig. 3: Methodology to assess video delivery.

All the wireless nodes run Ubuntu 10.04 on top of a Linux kernel 2.6.36. On the data nodes we use the official open-source `ath9k` driver for controlling the Atheros

cards; on the video nodes instead we adopt versions of the opensource `b43` driver and `OpenFirmware` firmware that we slightly modified for enabling the GATS mechanisms on the Broadcom cards. This custom firmware provides us with detailed statistics, that we used to confirm that we never observed packet drops at transmission or reception queues. Modifications are similar to what presented in [4] with minor performance improvements.⁷

The methodology works as follows. We use `ffmpeg` [31] to generate the reference video samples from the original uncompressed files. Then, we select one of the encoded videos, and the APM node streams it over the WLAN using a `GStreamer` [32] pipeline, which transforms the content into a packetised stream (`h264parse` element). Then, it chunks and encapsulates the stream into RTP packets (`rtph264pay` element), and sends those packets through a UDP socket to a multicast group using a given GATS scheme. During this transmission, each receiver use `tcpdump` [33] to collect `pcap` traces of the transmitted streams, which are stored in a RAM disk for performance reasons (i.e., to avoid losing data because of the SD card storage speed).

Once the streaming is completed, the trace files are uploaded to a separate computation server and processed offline to reconstruct the video from each received trace. Trace processing also uses `GStreamer` and basically follows the “reverse” procedure described above. It takes the stream configuration (`caps`, in `GStreamer` terminology) collected at the APM node for each experiment, and the `pcap` traces as input. Each stream is decapsulated, decoded and stored as a file of raw YUV frames. Such procedure includes a `GStreamer` element (`rtppjitterbuffer`) that simulates an initial reception buffer of 2 s.

To perform a full-reference computation of the video quality, we need that the reconstructed video consists of the same number of frames than the reference video, which may not happen if some packets are lost. To prevent this, we attach to each video an additional blank GOP (Group of Pictures) at the beginning. Then we artificially recover such a blank sequence at every receiver if it was found missing. From that point on, a `GStreamer` `videorate` element is capable of automatically maintaining the required frame rate by duplicating frames if losses occur (see [34] for further discussion on this issue), producing in this way a consistent reconstructed video with the expected number of frames after removing the initial blank GOP.

The selected full-reference metric is VMAF (Video Multi-Method Assessment Fusion), a recent perceptual video quality assessment algorithm developed by Netflix with a publicly available tool⁸ to compute this metric. The server computes the VMAF values by comparing the received video and the encoded video (i.e., not the source video, see Fig. 3). We decided this in order to have an adequate measurement of the performance degradation due to the wireless transmission, so differences in the degradation due to the encoding do not affect the VMAF values. Finally, we also used `Video Tester` [34] to extract other relevant

ID #	Description	Encoded size	VMAF
SRC 1	Red Kayak	5.45 MB	73.7
SRC 2	Ode to the West Wind	5.45 MB	92.3
SRC 3	Double End Bag	5.65 MB	96.7
SRC 4	Go Football	5.03 MB	87.3
SRC 5	Mr. Fins, Segment 2	4.80 MB	96.6
SRC 6	Halftime Music at Game	5.37 MB	91.1
SRC 7	Burn Close-up	5.40 MB	75.1
SRC 8	Zoom Out Showing Bottles	5.15 MB	96.2
SRC 9	Scarlet Oak Easy	5.20 MB	91.4

TABLE 1: Videos chosen for our measurements.

parameters such as the *packet skew*, or statistics about the type of frame lost.

3.3 Evaluation of the video quality

One of the key features of our evaluation is the use of a QoE-driven approach, which requires the use of real videos for streaming, instead of synthetic traffic generators, and an evaluation based on perceived quality, instead of metrics such as delay or packet loss ratio.

For the case of input videos, we selected a subset of the videos used by the Video Quality Experts Group (VQEG) and made available by the Ghent University-IBBT and the National Telecommunications and Information Administration (NTIA) at the *consumer digital video library*.⁹ More specifically, we downloaded the uncompressed original files (HRC 00 version) of the videos SRC 1–9 from the experiment `vqeghd1`, listed in Table 1. All of them are 10 s videos at 30 fps, with a HDTV 1080p (1920 × 1080, progressive) resolution and scanning format, and an approx. size of 1 GB.

Prior to their transmission over the WLAN, we compressed the videos in H.264 format with the `ffmpeg` tool following the `ffmpeg`’s encoding guide for streaming sites.¹⁰ We set up the `libx264` encoder with a “slow” preset, a target rate of 4 Mbps, and a GOP of 2 s (with a length of 60 frames, and a distance of 4 frames between anchors). With this choice, the encoded stream will provide to the receivers a complete video-frame – called “I” frame – every 2 s: if its transmission gets damaged, then the video would not recover until the next “I” frame, as the intermediate data – 59 “B” and “P” frames, with the pattern [BBBBP] – encodes only differences between consecutive “I” frames. In general, such losses have a great impact on the video quality. However, reducing the GOP requires either to increase the target rate to maintain the same quality, or to reduce the quality maintaining the same data rate. The video sizes following this initial trade-off choice are provided in Table 1, which illustrates notable compression ratios thanks to the use of lossy compression techniques. This brings up the issue of quality degradation and our QoE-evaluation, which we address next.

Given the vast amount of videos to be received and analysed, the use of a subjective Mean Opinion Score (MOS) evaluation is impractical (e.g., only for the experiments in Section 5, we process 15k videos). Because of this, we decided to assess the quality of the reception using Netflix’s VMAF algorithm to estimate video quality, which provides

7. The source code of the implementation is available at <http://netweb.ing.unibs.it/~openfirmwf>.

8. <https://github.com/Netflix/vmaf>

9. <http://www.cdv1.org>

10. <https://trac.ffmpeg.org/wiki/EncodingForStreamingSites>

state-of-the-art correlation with human perception [35], [36]. The VMAF is a full-reference metric that provides values between 0 (worst) and 100 (best). We note that, in general, QoE includes other parameters apart from video quality (e.g., start-up time, number and length of video stalls). However, given that in our setting the start-up time is very low (as we confirm later, delays are very small) and that the encoding ensures a constant frame rate (so frame losses result in frame duplication), throughout the paper we will use the terms “video quality” and “QoE” interchangeably.

We provide also in Table 1 the resulting quality of the videos after the encoding with the `ffmpeg` tool. As somehow expected, depending on the characteristics of the video, the resulting VMAF value varies (e.g., SRC 8 is more static than SRC 1), but in all cases they are relatively high, thus confirming the good quality of the encoded videos (we also confirmed this via visual inspection).

3.4 Performance of legacy multicast

We follow the previously described methodology to assess the quality of video delivery under the best radio conditions when using the legacy multicast service with the most robust MCS available, i.e., 6 Mbps. To this aim, we stream each of the nine videos over the WLAN using channel 14, which we assessed before as completely free from interference, and repeat the streaming three times. We run experiments during night-time to avoid (weak) cross-channel interference (e.g., from channels 11, 12, 13). Then, we sort stations by decreasing order of the received quality (considering the nine videos), and provide the resulting performance in Fig. 4, where we use one box-and-whisker plot per station to represent the 27 VMAF values (9 videos, 3 experiments per video).

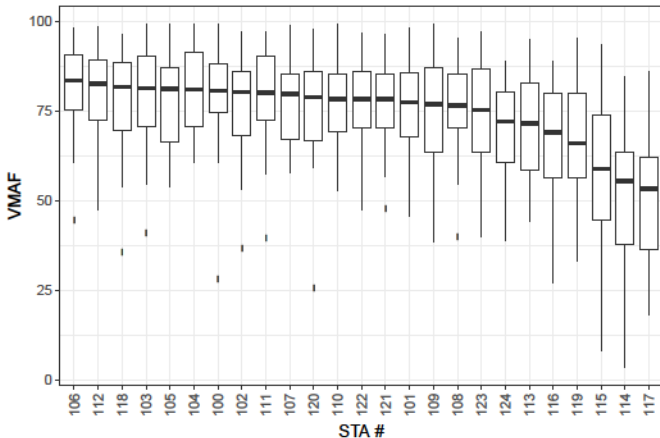


Fig. 4: VMAF performance for the legacy multicast at 6 Mbps, ideal conditions.

Fig. 4 illustrates that, even under these ideal conditions, the performance of the legacy multicast service is mediocre. Indeed, while for some stations the median of the VMAF is above e.g. 75, this is not guaranteed for all stations nor for all experiments, with a number of VMAF values falling well below 50. It can be seen that most of the worst performing stations are located in the room farthest away from the APM node (room 11 in Fig. 2), while some of the

best performing are located in its close proximity. However, given the complexity of wireless propagation, we do not attempt to find a relationship between distance from the APM node and QoE.

3.5 On the type and impact of losses

We next analyse the reasons behind the degraded video quality of the legacy service. Firstly, we confirmed via the statistics from the custom firmware that there are no packet drops at the transmission or reception queues, and therefore the degraded QoE can only be caused by (i) losses due to interference, or (ii) overly large delays that result in packet drops by the `videorate` element, which guarantees a constant video frame rate. To analyse whether the second happens, we compute the *skew* for all video packets, the skew being defined as the “RTP timestamp” minus the “arrival time,” which is zero for the first packet (to normalize) and negative in case a frame is delayed. As we found that minimum skew is approx. 100 ms (i.e., well below the reception buffer of 2 s), we discard the delay as a reason for degraded performance. In fact, throughout all experiments in this paper, we never found either losses at transmission or reception queues, or skews larger than a hundred ms, and therefore the only reason for degraded QoE are lost frames.

We then proceed to analyse the relation between packet losses and QoE. To this end, we analyse the traces from the same experiment of Fig. 4, and we compute for each station and video the % of packets that were correctly delivered, and we count the number of I-frames that were missing, as this has an impact on the perceived video quality [37], [38]. We then plot the VMAF vs. the delivery rate in Fig. 5, conditioned to the number of missing I-frames (each number with a different color). We make the following main observations from the results:

- Despite the mediocre performance observed, the delivery rates are very high—well above 99% for most of the cases. However, there is a notable span in terms of received video quality, filling the whole range of VMAF values.
- Furthermore, despite the positive correlation between reception rate and VMAF, this is rather small (the correlation coefficient R^2 is 0.3) there are many cases in which a similar delivery rate results in very different QoE. For instance, for reception rates around 99.5%, the VMAF values range from 25 to 100, which illustrates that different videos show different sensitivity to losses.
- Finally, we confirm that, in general, the higher the number of missing I-frames, the worse the video quality. However, we also note that given a number of missing I frames and similar loss rates, there are also notable differences in terms of the video quality (e.g., for no I frames lost, VMAF varies between 70 and 100 when the delivery rate is approx. 99.8%, for one I frame lost, VMAF varies between 50 and 90 when the delivery rate is approx. 99.5%).

The above confirms the need to perform QoE-driven evaluation of performance when considering the case of video service, as delivery rates constitute an imprecise indicator of video quality (at least for relatively high VMAF

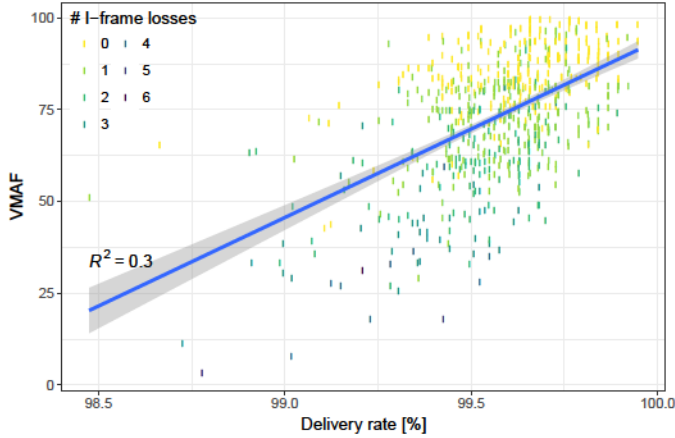


Fig. 5: VMAF vs. delivery rate using the legacy multicast at 6 Mbps.

values), even when information about the type and number of frames being lost is available. Furthermore, it illustrates the need to analyse performance building on a set of videos with different characteristics, given the different sensitivity to losses observed. In what follows, we use this VMAF-based evaluation to thoroughly evaluate the performance of GATS, starting from near-ideal scenarios and ending with a performance evaluation in the wild.

4 CONTROLLED SCENARIOS

We first analyse the performance of the GATS schemes in controlled scenarios. To this aim, we run our experiments night time and in channel 14, which is practically unoccupied throughout the experiments as we confirmed via frequent sniffing between experiments. Initially, we consider all possible MCS available for each GATS mechanism, although those schemes with many retransmissions (e.g., UR with 3 attempts per packet) cannot be used with overly low MCS, as this causes queue drops. More specifically, we considered the following GATS schemes and configurations:

- For the case of UR: (i) $R=1$, we run experiments with all MCS; (ii) $R=2$, with MCS higher or equal to 12 Mbps; and (iii) $R=3$, with MCS higher or equal to 36 Mbps. We do not consider $R>3$ values as we confirmed that they do not help to improve the quality of the video.
- For the case of BA, we run experiments for all MCS available. For this scheme, the only configuration parameter is the burst length N , which we decided to fix to 32 following our previous experimental results in [4], where this value resulted in the best performance (note that, as N decreases, the signalling load increases).

For each scheme and configuration, we streamed each of the nine considered videos, capturing the received packets at the 25 client stations and then computing the corresponding VMAF. As we are dealing with a relatively large number of values per video, we decide to use a summary statistic to represent the guaranteed QoE for a given video. For the sake of robustness, instead of e.g. the average or minimum

value, we decide to use a percentile. More specifically, our estimation of QoE from now on is given by the 5%-percentile of the set \mathbb{S}_j :

$$\text{Guaranteed VMAF} = Q_5(\mathbb{S}_j) \quad (2)$$

where $Q_p()$ represents the function to compute the p -percentile from a given set of values.

4.1 Ideal conditions

We start our analysis with interference, running the experiments during night-time in channel 14. We first consider the case of the UR scheme, with the resulting performance being illustrated in the three leftmost subfigures in Fig. 6, for the cases of one (UR=1), two (UR=2) and three (UR=3) transmissions attempts per L2 packet, respectively. Throughout this section, in order to have more robust measurements of the QoE guarantee, we repeat each experiment three times, and compute the percentile across experiments.

We first consider the case of UR=1, i.e., the leftmost subplot. From the figure, it is clear that the use of one transmission attempt per video packet, even under these ideal conditions, results in a bad performance for any MCS. Indeed, in all cases the VMAF is well below 75, so a high-quality video delivery is not guaranteed. It is also worth noting that, as seen in the previous sections, there are some videos (e.g., SSRC 7, SSRC 9) that show consistently lower values of VMAF than the others, resulting more sensitive to the (almost negligible) packet losses. Furthermore, although the use of a robust MCS helps to improve QoE performance (there is a clear difference in terms of performance between 54 Mbps and 6 Mbps), still the guaranteed performance is not adequate for a high-quality video streaming.

We next consider the case of UR=2, i.e., two transmission attempts per L2 packet at the configured MCS. As compared against the previous case, there is a notable improvement in performance for those MCS smaller than 48 Mbps, with the VMAF being higher than 75 in practically all MCS values between 6 Mbps and 36 Mbps (for the videos considered, this results in *watchable* video experience, as we confirmed for some selected experiments). For the cases of higher MCS, the performance is practically the same as in the case of UR=1, and therefore the QoE is very poor.

Finally, the use of three transmission attempts per L2 packet (UR=3) does not result in an appreciable performance improvement. In fact, apart from the 18 Mbps case, where video 1 suffered from notable losses, the VMAF performance is practically the same as in the UR=2 case. Given the additional transmission attempt, which results in a higher consumption of resources (wireless medium, energy), we conclude that for this ideal, interference-free scenario, there is no need to go beyond UR=2 transmission attempts. Furthermore, as performance is very similar across the MCS between 6 Mbps and 36 Mbps, it seems reasonable to select 36 Mbps as the MCS of choice, given that it results in the most efficient use of resources.

We next consider the BA scheme, with the results being depicted in the rightmost subfigure of Fig. 6. As compared to the UR scheme, here performance shows a much smaller variability for lower MCS values. In fact, for all MCS values smaller than 36 Mbps, stations are guaranteed a practically

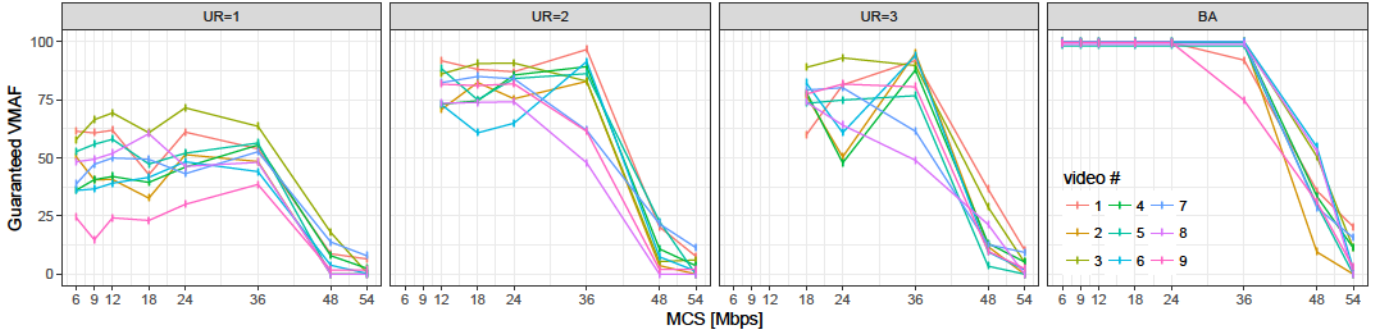


Fig. 6: Performance of the UR and BA schemes under ideal conditions.

“perfect” reception of all the videos, while for the case of 36 Mbps the guaranteed VMAF is also extremely high for all but two videos (which can also be watched with good QoE). For higher MCS, the performance is severely impaired, despite the fact that, with this scheme, each packet could be retransmitted up to a maximum of four times. These results confirm how challenging is to provide an excellent QoE performance with video over multicast for a relatively large set of receivers, as missing a single packet can have a notable impact on performance.

Following these results, it is clear that under ideal conditions, the BA scheme with a moderate MCS guarantees an excellent performance to all video receivers, thanks to the use of ARQ, while in contrast the performance of a properly configured UR is good at most. We note, though, that before deriving any conclusion about the “optimality” of these schemes, there are two (related) key issues that have to be considered, namely: the performance when there is interference, and the resources left by a scheme for non-video traffic. In what follows we address these issues.

4.2 Controlled interference

To analyse the performance when there is controlled activity in the wireless channel, we setup a different BSS with the Access Point and the station described in Section 3 (namely, APD and STAD in Fig. 2), using the standard DCF channel access. This BSS is configured to run on the same channel as the one for multicast transmission, i.e., in-band interference, and we generate interference at a rate R_i (which we will vary across experiments) by sending 1500 B UDP unicast traffic from the STAD to the APD node using `iperf`. To prevent the impact of the binary exponential backoff and the rate adaptation scheme on performance, we configure the wireless interface to transmit each packet only once at 54 Mbps.

To keep the number of experiments to a reasonable number, we first limit the number of schemes under consideration by discarding those that failed to provide an adequate performance under ideal conditions. In this way, following the results from the previous section, we do not consider anymore the case of UR=1, as its QoE performance is poor. Furthermore, we also do not consider the cases of 48 Mbps and 54 Mbps MCS, given that they result in poor performance as well. Out of the remaining configurations that provide an adequate performance (e.g., VMAF > 50), we restrict ourselves to the following schemes:

- For UR=2, we select the highest and lowest MCS that result in good performance, namely, 12 Mbps and 36 Mbps.
- For UR=3, following the same reasons, we select the pair MCS={18 Mbps, 36 Mbps}.
- Finally, for BA we select the lowest, highest and one intermediate value of MCS that achieved good performance, namely, MCS={6 Mbps, 12 Mbps, 36 Mbps}.

Our aim is to understand the impact of interference on performance. For each of the schemes and videos considered, we performed a sweep on the interference generation rate R_i , from 0 to the maximum achievable rate, which depends on the MCS configuration of the video transmissions (due to the *performance anomaly*). For each value of R_i we compute two figures: (i) the video delivery performance in terms of the guaranteed VMAF, as in the previous section, and (ii) the achieved data rate at the APD node, out of the generated rate R_i . This achieved rate, or goodput, will serve to measure the efficiency of a given multicast scheme, as quantifies the amount of resources available for non-video traffic.

The resulting performance for the UR scheme, in terms of VMAF to interference goodput, is depicted in Fig. 7. The results show that, in general, QoE rapidly degrades as soon as there is some other traffic in the channel, and that for this scheme it is not possible to have a configuration with good QoE for video and moderate data goodput. We next discuss in detail the obtained performance for the four considered configurations.¹¹

For the case of UR=2 at 36 Mbps (leftmost subplot), performance rapidly degrades with goodput: while with no interference (zero goodput) the QoE is adequate, the moment there is some interference (approx. 2 Mbps), the QoE significantly degrades for all videos, and rapidly goes to values below 0.25 for goodput values above 5 Mbps until the maximum value of R_i . For the case of UR=2 at 12 Mbps, which is a more robust scheme (subplot on the right to the previous one), in general the performance is slightly better, probably due to the *capture effect* (when being involved in a collision, the video packet is successfully received). Despite this advantage, the behaviour is still very similar:

11. We note that we are running again, for $R_i = 0$, some of the same experiments that we ran in the previous sections. Given the stochastic nature of losses, this causes minor variations on the Guaranteed VMAF.

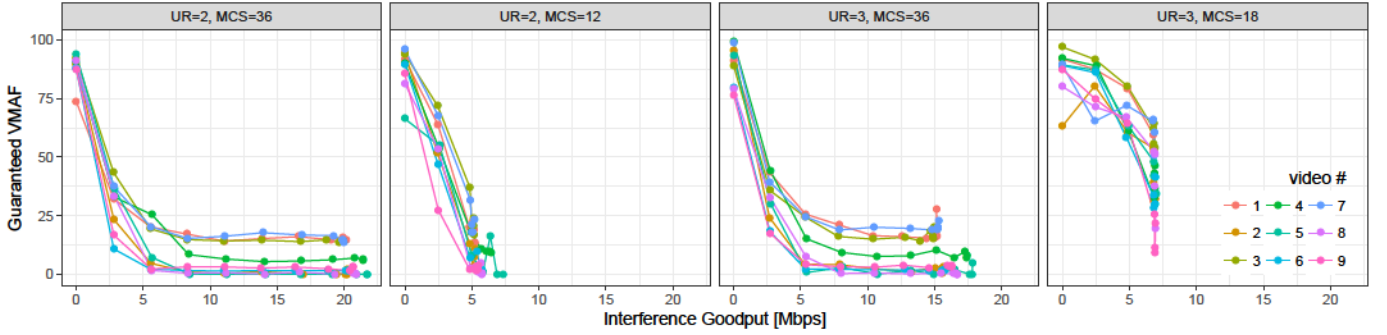


Fig. 7: Performance of UR under controlled interference.

for no interference the performance is good, but the moment data traffic is present, videos become unwatchable. Furthermore, given the lower MCS used in this case, due to the performance anomaly the maximum achievable goodput is smaller as compared against the previous case (approx. 6 Mbps vs. 20 Mbps).

We next consider the case of UR=3 at 36 Mbps (the third subplot). As compared to the first UR=2 case, the QoE is on average better, but like in the previous cases, the received videos are unwatchable when interference kicks in. Furthermore, because of the extra transmission attempt per packet, now in this case the maximum achievable goodput is smaller than in the case of UR=2 at 36 Mbps (approx. 17 Mbps vs. 20 Mbps). Finally, with UR=3 the use of the more robust MCS=18 Mbps (fourth subplot) does help to notably improve performance, with the quality of the videos being moderate even for a goodput of 5 Mbps. For higher interference rates, the quality of the majority of the received videos is bad, and the maximum achievable goodput in this case is approx. 7.5 Mbps.

Following these results, we conclude that the UR scheme might provide a relatively simple and adequate delivery of multicast video to a large set of receivers, but only in scenarios with no other traffic activity (UR=2), or very small interference (UR=3).

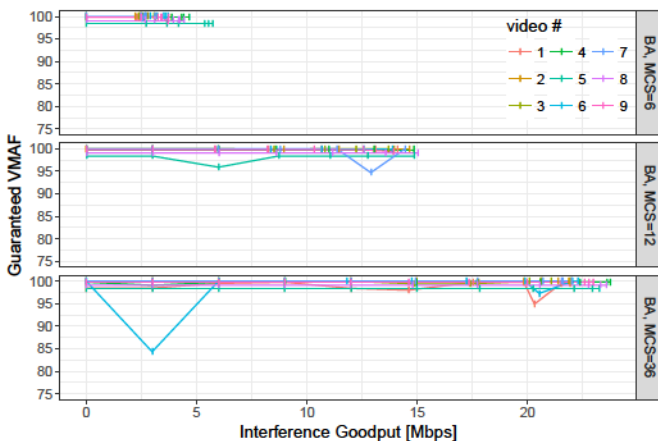


Fig. 8: Performance of BA under controlled interference.

We next analyse the obtained performance with the BA scheme, which we illustrate in Fig. 8. For this scheme, in all MCS considered the QoE is substantially higher than in

the previous cases, and practically insensitive to the interference. It should be noted, though, that the MCS affects the maximum achievable interference goodput, due to the *performance anomaly* effect [2]. For the case of the MCS=6 Mbps (top subplot), the guaranteed VMAF is practically 100 for all the videos. The price to pay for this performance is, as expected, notably small goodput values, which are bounded to approx. 5 Mbps. For the case of MCS=12 Mbps, the QoE is also maximum, while the maximum achievable goodput values are almost tripled. Finally, the use of MCS=36 Mbps results in the maximum achievable goodput figures, although in this case there is a small decrease in performance for video 6 when interference is around 3 Mbps (again, due to randomness).

Following these results for UDP controlled interference, we conclude that the BA scheme is able to guarantee a high QoE to all intended receivers, regardless of the amount of interference and without the need to use overly robust MCS. We next analyse if these conclusions hold under TCP traffic.

4.3 TCP traffic

To analyse the performance under TCP traffic, we configure the data station to generate a TCP flow towards the APD node at the maximum achievable rate. This flow starts 10 s before the transmission of the video, and ends after it. In contrast to the UDP case, we now configure the wireless interface to the usual transmission settings (i.e., up to 7 retransmission attempts), to reduce the impact of L2 losses on TCP's congestion control. We proceed as in the previous case, streaming the 9 considered videos and computing on the one hand the guaranteed VMAF, and on the other hand the goodput obtained by TCP. We consider the same GATS schemes and configuration as in the previous case, plus some additional representative ones.

We depict the resulting values in Fig. 9, in which experiments corresponding to the same GATS configuration are enclosed in the same grey box and denoted with the corresponding label (note that there is some overlap between boxes). We start by considering the UR schemes which, in general, result in a trade-off between the Guaranteed VMAF and the TCP goodput, as there is a negative correlation between these two figures: the higher the number of retries, or the more robust the MCS, the better the QoE but the less resources are left available for data traffic. However, in all UR configurations the video quality is low: for all UR=2

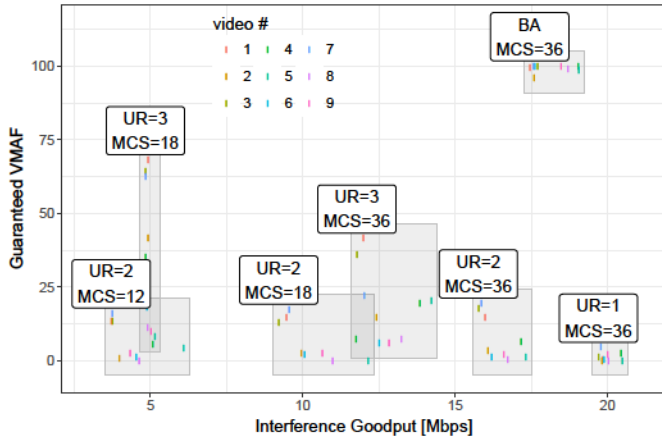


Fig. 9: Performance under TCP traffic.

cases the guaranteed VMAF is at most 25, and well below this value for most of the videos; the use of UR=3, like in the case of controlled interference, helps to improve the VMAF but still the majority of the videos are not delivered with an adequate performance, and only with the 18 Mbps MCS some guaranteed values are above 50.

In contrast to the mediocre performance of UR under TCP interference, the BA scheme is able to guarantee an excellent video delivery, while also providing a very high TCP throughput. Indeed, the goodput figures are quite similar to the ones obtained with the {UR=1, 36 Mbps} configuration, which is represented in the figure for comparison as this scheme is the one resulting in the least occupation of the medium at 36 Mbps (with a guaranteed VMAF of approx. zero). According to these results, the BA scheme is able to guarantee VMAF at a cost of approx. 2 Mbps in terms of TCP goodput, as compared to the most efficient scheme (note that we do not represent other MCS values for BA, as according to Fig. 8, they lead to the same excellent QoE performance but achieve a lower goodput).

The conclusions in the case of TCP controlled interference are similar to those we observed for UDP, in particular when the amount of interference was relatively high: the UR schemes are not able to guarantee an adequate VMAF to all receivers, despite consuming a significant amount of wireless resources, while the BA scheme is able to maximise performance for both variables, with its “hidden cost” being the relative complexity of the scheme. Given that in a real deployment the interference may change over time, in Section 5 we address the performance of the schemes under uncontrolled interference.

4.4 Impact of EDCA differentiation

One of the objectives of the EDCA mechanism introduced by the 802.11e amendment is to provide service differentiation (and, to some extent, QoS) by configuring the contention parameters of the different access categories with specific values. To assess the impact of this differentiation, we repeat a subset of the previous experiments, but now with the 802.11 interfaces of the data stations configured with the recommended parameters for Best Effort category, which results in a lower priority when accessing the channel

as compared to the Video category (the AIFS parameter is one empty slot longer in the former than in the latter).

For simplicity, we consider the following GATS mechanisms: BA, UR = {1, 2, 3}, and MCS = {18, 36} Mbps. Also, for the case of UDP interference we consider the same generation rates as before, which results in approx. 2000 experiments for UDP, and 200 experiments for TCP, where in each experiment we have a pair of VMAF and goodput values. Then, for each experiment we determine the difference in performance between the previous case and the EDCA differentiation, by computing (i) on the one hand, the VMAF difference (which we expect positive, given the prioritisation of video traffic):

$$\text{VMAF difference} = \text{VMAF(EDCA)} - \text{VMAF(DCF)} \quad (3)$$

and (ii) on the other hand, the goodput loss (we expect a drop in data performance, for the same reasons):

$$\text{Goodput loss} = \frac{\text{Goodput(DCF)} - \text{Goodput(EDCA)}}{\text{Goodput(DCF)}} \quad (4)$$

In order to formally determine if there is any change in performance thanks to the use of the EDCA differentiation, we perform a pair of Student’s *t*-tests over the collected sets of differences and losses, with the null hypotheses being that the mean of the population is zero, i.e., that there is no significant difference in VMAF values, nor any significant difference in Goodput loss. The results, with the corresponding confidence intervals, are depicted in Fig. 10 for the cases of TCP (top) and UDP (bottom).

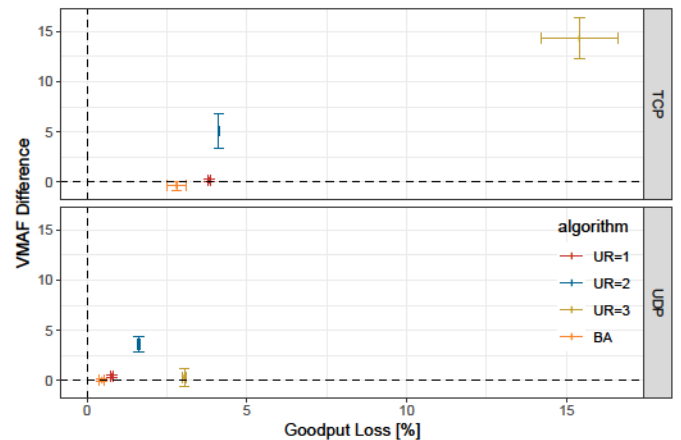


Fig. 10: Impact of EDCA differentiation.

The results from this analysis show that, in general, the use of EDCA has very little impact on performance. For the case of the UDP experiments, there are practically no differences in performance in terms of VMAF improvement, with only the UR=2 case not overlapping with the 0 line (note that the absolute average gain is quite small). Considering goodput losses, these are as well very small and well below 5%. Despite one could expect that for the UDP experiments the impact of EDCA differentiation is moderate, as the data station is not saturated in many cases, it is remarkable such small differentiation.

We next consider the case of the TCP experiments, in which the data station is saturated and therefore we expect

higher EDCA differentiation. Here, as the top of Fig. 10 illustrates, the differences in performance are more notable, but for the case of BA, in which there is a small goodput loss (approx. 3%). For the UR cases, there are small improvements in VMAF in exchange for some goodput losses, with the maximum improvement being approx. 15 with a corresponding throughput loss of approx. 15%. Given that this improvement in QoS is not very significant, and fails to result in a good QoE, we conclude that the EDCA differentiation is of little interest in our scenario.

5 PERFORMANCE IN THE WILD

We finalise our study by performing a series of experiments under non-controlled interference, i.e., the one caused by the students and faculty using the WLANs near or inside our lab scenario. To this end, we select on the one hand a set \mathbb{C} of mechanisms comprising the legacy scheme ($\{\text{UR}=1, 6 \text{ Mbps}\}$) plus a subset of the best performing mechanisms and configurations from the previous results, namely, three UR configurations ($\{\text{UR}=2, 18 \text{ Mbps}\}$, $\{\text{UR}=3, 18 \text{ Mbps}\}$, and $\{\text{UR}=3, 36 \text{ Mbps}\}$), and the $\{\text{BA}, 36 \text{ Mbps}\}$ case, and on the other hand a subset of the available WLAN channels, $\{1, 6, 11 \text{ and } 14\}$.

We create a list of 20 experiments by repeating the five mechanisms in \mathbb{C} over the four channels, and we select randomly one of the nine available videos for each experiment. We then randomise the list before starting the transmissions, after which we collect the traces at each receiver. We repeat this procedure in a loop during a working day starting at 8 AM and finishing at approx. 1 AM of the following day. For each video trace at each station, we compute the corresponding VMAF, and represent the resulting values with dots in Fig. 11 (there are approx. 15k dots), where we use a different color for each different channel. Lines represent rolling averages across stations over a period of around 2 hours. In this section, given the variability of the results, for ease of visualisation we decided to represent these variables instead of the guaranteed VMAF.

As results show, the legacy scheme ($\{\text{UR}=1, 6 \text{ Mbps}\}$) cannot guarantee an adequate QoE during the majority of the time and channels. Indeed, for the most populated channels (1, 6 and 11), the average performance is well below 0.5 in practically all experiments. Despite in channel 14 the performance is much better, still only during the late evening (from 19 h on) there is a notable improvement of the VMAF, caused by a reduction of the interference in the WLAN. These results confirm that the use of the legacy scheme, even with the most robust MCS available, leads to a poor multicast video experience.

We next consider the other UR cases, which in general improve performance as compared to the legacy scheme but also fail to guarantee an excellent performance in all experiments. Comparing the cases of $\{\text{UR}=2, 18 \text{ Mbps}\}$ and $\{\text{UR}=3, 18 \text{ Mbps}\}$, it is clear that the use of three transmission attempts supports a noticeable improvement of QoE before 19 h (see e.g. the performance in channel 1, among the most populated in our setting). This benefit from the extra transmission attempt seems to be less relevant from 21 h on, i.e., when the laboratory is empty. During this late period of the day, both schemes provide an adequate

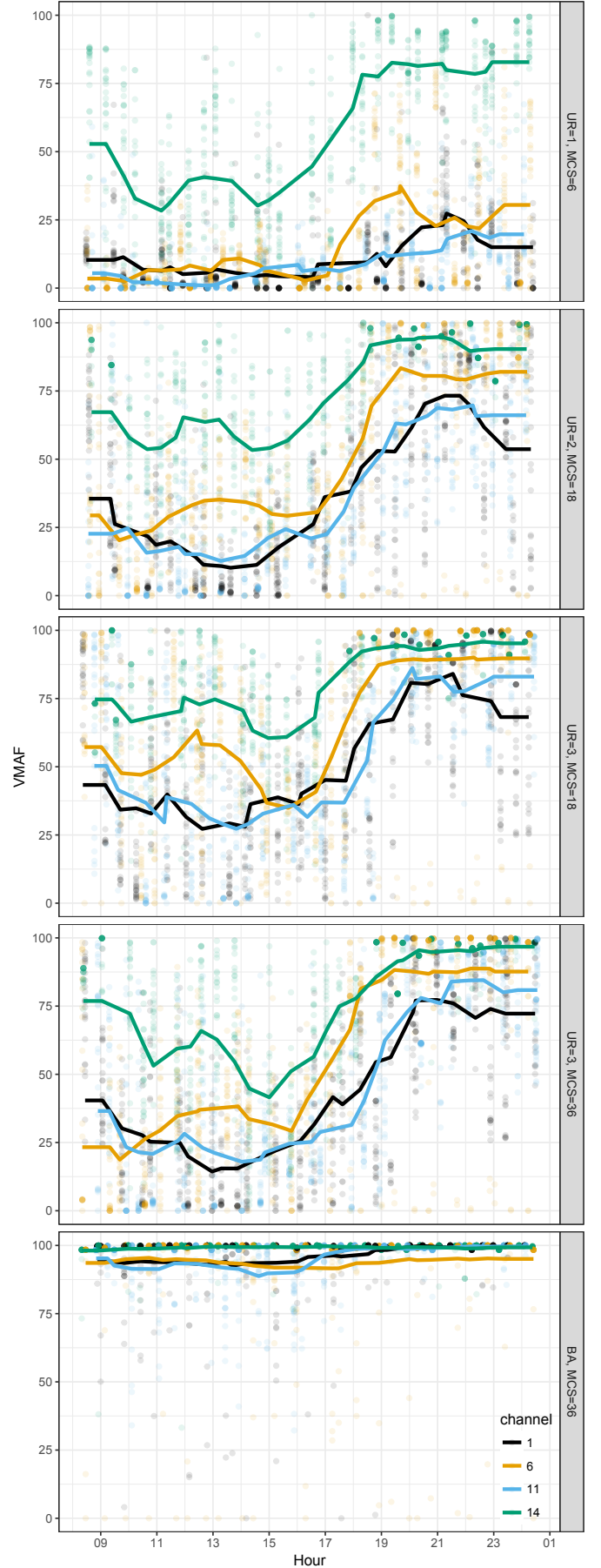


Fig. 11: VMAF performance in the wild.

QoE for channels {1, 6, 11} and an excellent video quality for channel 14. A very similar comparison can be made considering the cases of {UR=3, 36 Mbps} and {UR=3, 18 Mbps}: the use of a more robust MCS provides some advantages when the interference is high, but these benefits are less evident the moment the activity in the other WLANs decreases. Furthermore, despite the performance is improved, the resulting video quality is mediocre at most.

Finally, the results from the BA scheme are practically constant during the entire experiment: in all cases, the obtained performance is excellent, with the VMAF well above 80 and only a few stations failing to receive the video with such good quality. These cases, as we confirmed by analysing the traces and doing ad hoc experimentation, are caused by the mobility of users in the lab, which can temporarily destroy the link between the APM node and one of the stations. We finalize our study by analyzing the packet losses experienced when using the different mechanisms.

5.1 Analysis of the packet losses

To gain insight into the causes of the observed performance, we further analyzed the obtained traces to characterize the losses experienced by each mechanism. To this aim, we count for the different mechanisms the amount of missing RTP packets from each video transmission at each station: we then compute the experimental Cumulative Distribution Function (eCDF) that we illustrate in Fig. 12. Results¹² are two-fold: first, we confirm that video quality is extremely sensitive to losses, as the the large disparity of VMAF results already seen in Fig. 11 is not corresponded by the relatively small loss ratios (note that apart from the UR=1 case, losses are in general well below 5%); second, each mechanism is providing the qualitative expected performance, with BA achieving significantly smaller losses than the UR-based schemes. Results confirm that performance improves both when increasing the number of transmission attempts, i.e., from {UR=2, MCS=18} to {UR=3, MCS=18}, or when using a more robust MCS, i.e., from {UR=3, MCS=36} to {UR=3, MCS=18}. Finally, it is worth mentioning the relatively similar performance of {UR=2, MCS=18} and {UR=3, MCS=36}, something already observed in Fig. 9.

We also analyzed, for each mechanism, the distribution of the number of consecutive losses, to understand whether these happen sparsely or in bursts. Following this analysis, we observe that more than 95% of the losses appear in bursts of length 5 at most: this confirms that relatively long bursts of losses, that can be caused by the mobility of users in the lab mentioned above, appear very sporadically.

6 CONCLUSIONS, SCOPE AND FUTURE WORK

We have performed an extensive QoE-driven evaluation of the novel mechanisms for multicast made available by the 802.11aa amendment. We have designed a thoughtful methodology, building on a standardised set of videos, and assessed the performance of the different GATS mechanisms under different configurations and a variety of scenarios.

12. Throughout this section, for ease of representation we provide the results corresponding to channel 11, but the qualitative conclusions also apply to the other channels.

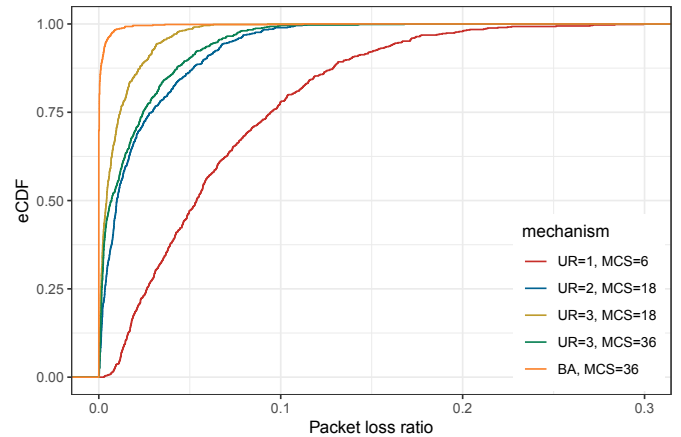


Fig. 12: Experimental CDF of loss ratios.

These scenarios have considered different assumptions such as ideal conditions, UDP and TCP controlled interference, or a 15 h performance evaluation in the wild. To the best of our knowledge, our results provide the first real-life, QoE driven evaluation of the performance of GATS, and illustrate their different trade-offs and the good properties of the BA scheme (once an adequate MCS is configured).

Scope: Our study is limited to the used encoding parameters, which have an impact on both the transmission pattern generated, and on the relation between losses and the perceived quality. It is also tied to the recovery scheme used by the considered mechanisms, i.e., frame re-transmission, instead of more advanced ones (based on, e.g., on forward error correction, network coding, or layered encoding). Furthermore, while channel conditions are not ideal (e.g., noise and in-band interference), we did not experience hidden nodes or other challenging scenarios. This motivates new lines of research, discussed next.

Future work: By releasing the source code produced and the tracefiles collected during the experiments, we enable other researchers to repeat or extend our methodology at a low implementation cost. For instance, one line of research could analyze the impact of video encoding on performance, the use of more advanced re-transmission schemes, or the design of cross-layer techniques to maximise QoE. It would be straightforward to derive configuration guidelines for other WLAN scenarios by repeating our methodology. Leveraging on these guidelines, one could design adaptive algorithms for multimedia Access Point, which change the mechanism for multicast streaming depending on the estimated conditions of the WLAN.

ACKNOWLEDGEMENTS

We thank Luca Cominardi and Pablo Salvador for their contributions on the initial implementation of GATS. We also thank the anonymous reviewers for their comments, which helped improve the final version of this manuscript. The work of F. Gringoli was partly supported by the European Commission (EC) in the framework of the H2020-ICT-2014-1 project WiSHFUL (Grant agreement no. 645274). The work of P. Serrano and I. Ucar were partially supported by the EC in the framework of H2020-ICT-2014-2 project

Flex5Gware (Grant agreement no. 671563), by the Spanish Ministry of Economy and Competitiveness through the 5G-City project (TEC2016-76795-C6-3-R), and by the Madrid Regional Government through the TIGRE5-CM program (S2013/ICE-2919).

REFERENCES

- [1] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 3: MAC Enhancements for Robust Audio Video Streaming*, IEEE Amendment 802.11aa, 2012.
- [2] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 2, March 2003, pp. 836–843 vol.2.
- [3] A. Banchs, A. De La Oliva, L. Eznarriaga, D. Kowalski, and P. Serrano, "Performance Analysis and Algorithm Selection for Reliable Multicast in IEEE 802.11aa Wireless LAN," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3875–3891, Oct 2014.
- [4] P. Salvador, L. Cominardi, F. Gringoli, and P. Serrano, "A First Implementation and Evaluation of the IEEE 802.11Aa Group Addressed Transmission Service," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 1, pp. 35–41, Dec. 2013.
- [5] J. M. Vella and S. Zammit, "A survey of multicasting over wireless access networks," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 718–753, Second 2013.
- [6] H. I. Zawia, R. Hassan, and D. P. Dahnil, "A survey of medium access mechanisms for providing robust audio video streaming in ieee 802.11aa standard," *IEEE Access*, vol. 6, pp. 27 690–27 705, 2018.
- [7] A. de la Oliva, P. Serrano, P. Salvador, and A. Banchs, "Performance evaluation of the ieee 802.11aa multicast mechanisms for video streaming," in *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, June 2013, pp. 1–9.
- [8] Y. Shin, M. Choi, J. Koo, and S. Choi, "Video multicast over wlans: Power saving and reliability perspectives," *IEEE Network*, vol. 27, no. 2, pp. 40–46, March 2013.
- [9] Q. Li, L. Jiao, and F. Y. Li, "Performance evaluation of the gcr block ack mechanism in ieee 802.11aa networks," in *European Wireless 2013; 19th European Wireless Conference*, April 2013, pp. 1–7.
- [10] T. Li, Q. Ni, T. T. Turletti, and Y. Xiao, "Performance analysis of the ieee 802.11e block ack scheme in a noisy channel," in *2nd International Conference on Broadband Networks, 2005.*, Oct 2005, pp. 511–517 Vol. 1.
- [11] K. Mansour, I. Jabri, and T. Ezzedine, "Ieee 802.11aa gats mechanisms performance study for very high throughput wlans," in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC 2018), Barcelona (Spain), Apr. 15-18, 2018*.
- [12] V. Gupta, Y. Bejerano, C. Gutterman, J. Ferragut, K. Guo, T. Nandagopal, and G. Zussman, "Light-weight feedback mechanism for wifi multicast to very large groups - experimental evaluation," *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3826–3840, Dec 2016.
- [13] R. Chandra, S. Karanth, T. Moscibroda, V. Navda, J. Padhye, R. Ramjee, and L. Ravindrananth, "Dircast: A practical and efficient wi-fi multicast system." *IEEE*, October 2009.
- [14] X. Wang, L. Wang, Y. Wang, Y. Zhang, and A. Yamada, "Supporting mac layer multicast in ieee 802.11n: Issues and solutions," in *2009 IEEE Wireless Communications and Networking Conference*, April 2009, pp. 1–6.
- [15] M. Choi, W. Sun, J. Koo, S. Choi, and K. G. Shin, "Reliable video multicast over wi-fi networks with coordinated multiple aps," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 424–432.
- [16] K. Mansour, I. Jabri, and T. Ezzedine, "New 802.11aa ack leader selection scheme for multicast qos improvement," in *2016 Wireless Days (WD)*, March 2016, pp. 1–6.
- [17] Y. Daldoul, T. Ahmed, and D. E. Meddour, "Ieee 802.11n aggregation performance study for the multicast," in *2011 IFIP Wireless Days (WD)*, Oct 2011, pp. 1–6.
- [18] H. Wang, W. T. Ooi, and M. C. Chan, "Jurcast: Joint user and rate allocation for video multicast over multiple aps," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, April 2016, pp. 1–9.
- [19] A. Ksentini, M. Naimi, and A. Gueroui, "Toward an improvement of h.264 video transmission over ieee 802.11e through a cross-layer architecture," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 107–114, Jan 2006.
- [20] W. S. Lim, D. W. Kim, and Y. J. Suh, "Design of efficient multicast protocol for ieee 802.11n wlans and cross-layer optimization for scalable video streaming," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 780–792, May 2012.
- [21] Y. Shin, S. Baek, S. Choi, and B. G. Lee, "Scalable video multicast using flexible multicast service over ieee 802.11 wlan," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.
- [22] K. C. J. Lin, W. L. Shen, C. C. Hsu, and C. F. Chou, "Quality-differentiated video multicast in multirate wireless networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 21–34, Jan 2013.
- [23] R. Muzaffar, E. Yanmaz, C. Bettstetter, and A. Cavallaro, "Application-layer rate-adaptive multicast video streaming over 802.11 for mobile devices," in *Proceedings of the 2016 ACM on Multimedia Conference*, ser. MM '16. New York, NY, USA: ACM, 2016, pp. 506–510. [Online]. Available: <http://doi.acm.org/10.1145/2964284.2967272>
- [24] J. Pokhrel, I. Paudel, B. Wehbi, A. Cavalli, and B. Jouaber, "Performance evaluation of video transmission over 802.11n wireless network: A mac layer perspective," in *2014 International Conference on Smart Communications in Network Technologies (SaCoNeT)*, June 2014, pp. 1–6.
- [25] G. Barreira and J. Ascenso, "Impact of the ieee 802.11n frame aggregation mechanisms on video streaming quality," in *SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks*, Sept 2012, pp. 1–5.
- [26] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, "Measuring video qoe from encrypted traffic," in *Proceedings of the 2016 ACM on Internet Measurement Conference*, ser. IMC '16. New York, NY, USA: ACM, 2016, pp. 513–526. [Online]. Available: <http://doi.acm.org/10.1145/2987443.2987459>
- [27] A. O. Adeyemi-Ejeye and S. D. Walker, "Ultra-high definition wireless video transmission using h.264 over 802.11n wlan: Challenges and performance evaluation," in *Proceedings of the 12th International Conference on Telecommunications*, June 2013, pp. 109–114.
- [28] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121 – 132, 2004.
- [29] A. Kamerman and G. Aben, "Throughput performance of wireless lans operating at 2.4 and 5 ghz," in *11th IEEE International Symposium on Personal Indoor and Mobile Radio Communications. PIMRC 2000. Proceedings (Cat. No.00TH8525)*, 2000.
- [30] S. Fiehe, J. Riihijärvi, and P. Mähönen, "Experimental study on performance of ieee 802.11n and impact of interferers on the 2.4 ghz ism band," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference (IWCMC '10)*. ACM, New York, NY, USA, 2010.
- [31] FFmpeg Developers, *ffmpeg tool*, 2017. [Online]. Available: <https://ffmpeg.org>
- [32] GStreamer Team, *GStreamer framework*, 2017. [Online]. Available: <https://gstreamer.freedesktop.org>
- [33] tcpdump developers, *tcpdump tool*, 2017. [Online]. Available: <http://www.tcpdump.org>
- [34] I. Ucar, J. Navarro-Ortiz, P. Ameigeiras, and J. M. Lopez-Soler, "Video tester: A multiple-metric framework for video quality assessment over ip networks," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, June 2012, pp. 1–5.
- [35] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward A Practical Perceptual Video Quality Metric," June 2016. [Online]. Available: <https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652>
- [36] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2017, pp. 1–2.
- [37] M. Mu, R. Gostner, A. Mauthe, G. Tyson, and F. Garcia, "Visibility of individual packet loss on h.264 encoded video stream: a user study on the impact of packet loss on perceived video quality," in *Proc. SPIE 7253, Multimedia Computing and Networking 2009*, 725302, 2009.

- [38] Q. Dai and R. Lehnert, "Impact of packet loss on the perceived video quality," in *2nd International Conference on Evolving Internet, Valencia, 2010*, pp. 206-209., 2010.



Francesco Gringoli (M'04, SM'17) received his Laurea degree in Telecommunications Engineering from the University of Padua (Italy) in 1998 and his Ph.D. degree in Information Engineering from the University of Brescia (Italy) in 2002. He is Associate Professor of Telecommunications at the Dept. of Information Engineering of the University of Brescia, Italy. He started the Open-FWWF Project in 2009.



Arturo Azcorra received his M.Sc. degree in Telecommunications Engineering from the Universidad Politécnica de Madrid in 1986 and his Ph.D. from the same university in 1989. In 1993, he obtained an M.B.A. from the Instituto de Empresa. He has a double appointment as Full Professor (with chair) at the Telematics Engineering Department of University Carlos III of Madrid and as Director of IMDEA Networks. Prof. Azcorra has coordinated the CONTENT and E-NEXT European Networks of Excellence, has served as a Program Committee Member in numerous international conferences and has published over 100 scientific papers in books, international journals and conferences.



Pablo Serrano (M'09, SM'16) received his degree in telecommunication engineering and his Ph.D. from Universidad Carlos III de Madrid (UC3M) in 2002 and 2006, respectively. He has been with the Telematics Department of UC3M since 2002, where he currently holds the position of associate professor. He has over 80 scientific papers in peer-reviewed international journal and conferences. He has served as guest editor for *Computer Networks*, and on the TPC of a number of conferences and workshops including IEEE INFOCOM, IEEE WoWMoM and IEEE Globecom.



Iñaki Ucar received his M.Sc.Eng. in Telecommunications Engineering and M.Sc. in Communications from Universidad Pública de Navarra (UPNA) in 2011 and 2013 respectively, and his M.Sc. and Ph.D. in Telematic Engineering from Universidad Carlos III de Madrid (UC3M) in 2014 and 2018 respectively. His work focuses on energy efficiency of wireless networks.



Nicoló Facchi is a postdoc research fellow at the Department of Information Engineering and Computer Science of the University of Trento, Italy. He received the Ph.D. in Telecommunication Engineering from the University of Brescia in 2016. His current research interests include real time video streaming and overlay management in P2P networks and performance evaluation of medium access control in Wireless LANs.