

Improving Content Delivery Efficiency through Multi-Layer Mobile Edge Adaptation

Jose Oscar Fajardo, Ianire Taboada, Fidel Liberal

University of the Basque Country (UPV/EHU)

{joseoscar.fajardo, ianire.taboada, fidel.liberal}@ehu.eus

This document is the “**accepted article**” version of the paper, which has been accepted by IEEE for publication in **IEEE Network** magazine, issue of Noviembre 2015 (feature topic: “Quality-of-Experience (QoE)-Aware Design in Next-Generation Wireless Networks”).

Copyright has been transferred to IEEE on 18th September 2015.

Full citation information:

Jose Oscar Fajardo, Ianire Taboada, Fidel Liberal, “Improving content delivery efficiency through multi-layer mobile edge adaptation”, IEEE Network, Volume29, Issue 6, pp. 40-46, Nov.-Dec. 2015.

Links to the published version:

DOI: [10.1109/MNET.2015.7340423](https://doi.org/10.1109/MNET.2015.7340423)

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7340423

The research leading to these results has received funding from:

- The European Union's H2020 programme, under the Grant Agreement H2020-ICT-671596 Research and Innovation Project SESAME - Small cEllS coordinAtion for Multi-tenancy and Edge services
- The Spanish Ministerio de Economía y Competitividad (MINECO) under grant TEC2013-46766-R: QoEverage - QoE- aware optimization mechanisms for next generation networks and services

Improving Content Delivery Efficiency through Multi-Layer Mobile Edge Adaptation

Jose Oscar Fajardo*, Ianire Taboada, Fidel Liberal

University of the Basque Country (UPV/EHU), Alameda Urquijo s/n, 48013 Bilbao, Spain
{joseoscar.fajardo, ianire.taboada, fidel.liberal}@ehu.eus

Abstract

This paper presents a novel architecture for optimizing the HTTP-based multimedia delivery in multi-user mobile networks. This proposal combines the usual client-driven dynamic adaptation scheme DASH-3GPP with network-assisted adaptation capabilities, in order to maximize the overall Quality of Experience. The foundation of this combined adaptation scheme is based on two state of the art technologies. On one hand, adaptive HTTP streaming with multi-layer encoding allows efficient media delivery and improves the experienced media quality in highly dynamic channels. Additionally, it enables the possibility to implement network-level adaptations for better coping with multi-user scenarios. On the other hand, mobile edge computing facilitates the deployment of mobile services close to the user. This approach brings new possibilities in modern and future mobile networks, such as close to zero delays and awareness of the radio status. The proposal in this paper introduces a novel element, denoted as Mobile Edge-DASH Adaptation Function, which combines all these advantages to support efficient media delivery in mobile multi-user scenarios. Furthermore, we evaluate the performance enhancements of this content- and user context-aware scheme through simulations of a mobile multimedia scenario.

Mobile multimedia traffic has been experiencing a dramatic increase in the last years, dominated by the explosion of media delivery through Dynamic Adaptive Streaming over HTTP (DASH). This approach, standardized by MPEG as ISO/IEC CD 23009-1 and adopted by 3GPP as 3GP-DASH, splits media contents in short media segments. These chunks can be made available at different quality versions, allowing users to switch between different quality representations from one interval to the following one. In this way, multimedia services are endowed with client-driven dynamic adaptation capabilities, which is a crucial feature for reacting to variable channel conditions.

DASH is usually deployed with single-layer video codecs such as H.264/Advanced Video Coding (AVC) and H.265/High Efficiency Video Coding (HEVC). However, there is an increasing interest in endowing DASH with the flexibility supported by multi-layer video codecs such as Scalable Video Coding (SVC). A number of recent studies analyze the potential performance enhancements introduced by DASH and SVC [1, 2, 3, 4].

Figure 1 illustrates the transmission fundamentals concerning single-layer (Fig. 1a) and multi-layer (Fig. 1b) DASH.

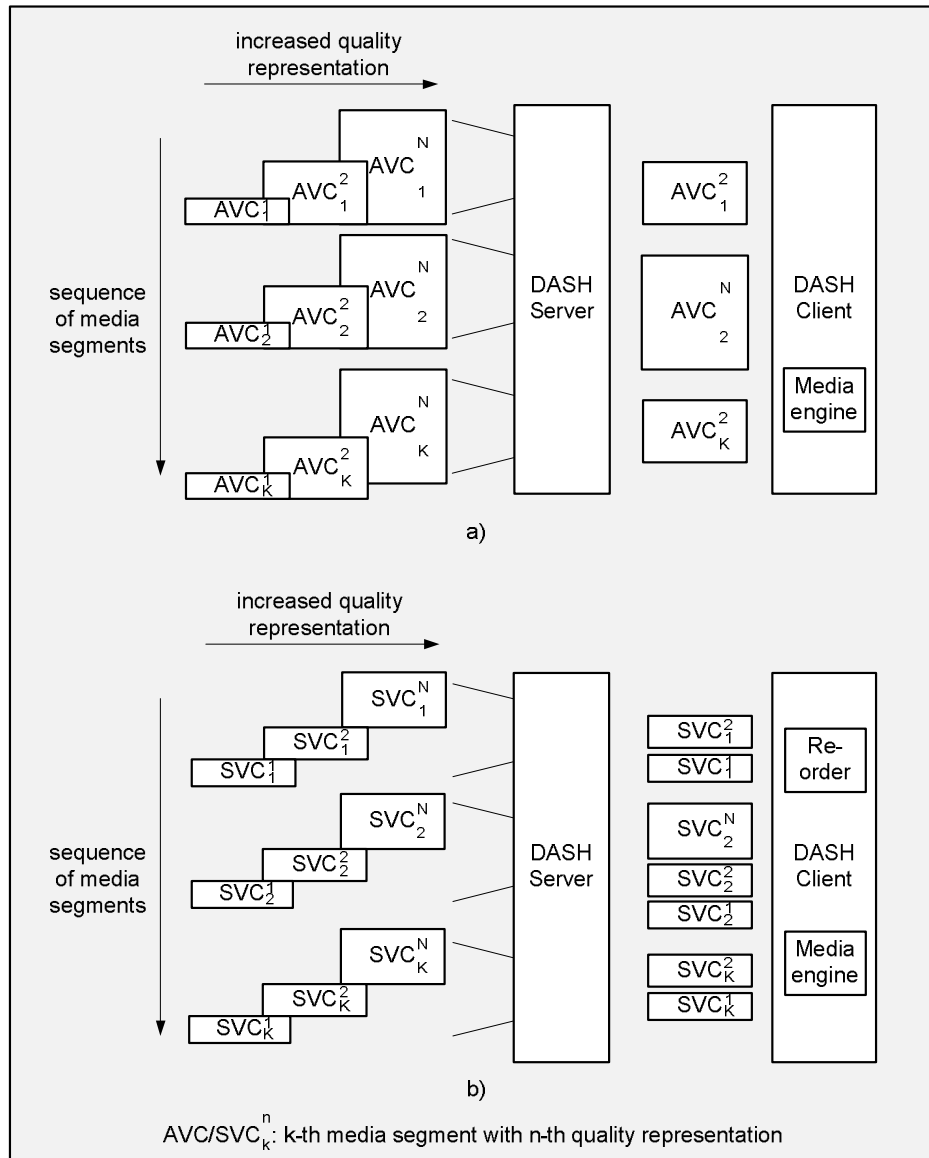


Figure 1. DASH transmission scheme: a) AVC; b) SVC.

In AVC-based DASH (DASH-AVC), media content is split in K media segments (AVC_k in Fig. 1a) available at N different representations (AVCⁿ in Fig. 1a), corresponding to different quality levels. Each combination of media segment and quality representation is described as a unique HTTP object in a Media Presentation Description (MPD) file, and can be independently retrieved through HTTP GET requests. Therefore, at each time slot, the user device requests a unique video representation.

SVC-based DASH (DASH-SVC) allows more flexible delivery schemes since different layers (base layer and enhancement layers) are split into different HTTP objects containing additive information. Each HTTP object in the MPD file represents a quality layer of a video segment. When client devices select the most suitable media representation (SVC¹-SVCⁿ objects in Fig.

1b) for a media segment (SVC_k in Fig. 1b), the different layers are transmitted over the network as standalone HTTP transactions.

Dynamic content delivery reacts to the specific network conditions based on either quick client-driven or network-driven decisions. However, as discussed through this paper, we find a series of limitations in the existing adaptation approaches. First, client-driven adaptations are individually driven and the combined effects of sharing radio resources are not fully considered. Second, network-driven adaptation elements lack of accurate radio channel awareness, which hinders the optimal delivery of media sessions. Finally, current radio network elements offer limited capacity to include Quality of Experience (QoE) awareness.

This paper focuses on the mobile network-assisted adaptation of media flows in the context of DASH-SVC, aimed at maximizing the overall QoE in scenarios where multiple mobile users are accessing multimedia content at the same time through a shared radio access. Among the different QoE influence factors [5, 6], in the context of this paper the QoE for each user is determined by the subjective quality associated to the achieved content bitrate and gauged according to [1] and [7].

In order to achieve this objective, we propose a novel architecture intended to exploit both the content awareness and the user context awareness. This proposal includes a network element, denoted as Mobile Edge-DASH Adaptation Function (ME-DAF), which is located within the Cloud Radio Access Network (C-RAN) [8] following the novel Mobile Edge Computing (MEC) principles [9]. ME-DAF integrates media caching and QoE-driven traffic conforming capabilities.

First, we overview different client- and network-driven DASH adaptation proposals. Next, we present the proposed architecture and introduce the ME-DAF element. Performance evaluation is carried out by simulating the behavior of LTE schedulers, client-driven adaptations and QoE-driven edge network-based adaptations in a mobile scenario taking into account real-world channel quality traces. Finally, we provide the conclusions to the paper.

Evolution of multimedia session adaptation approaches

Different proposals aim at optimizing the transmission of multimedia content over the Internet, with special interest in mobile environments. Figure 2 illustrates the evolution of network architectures supporting DASH-based media delivery, with different adaptation points available over the provision chain.

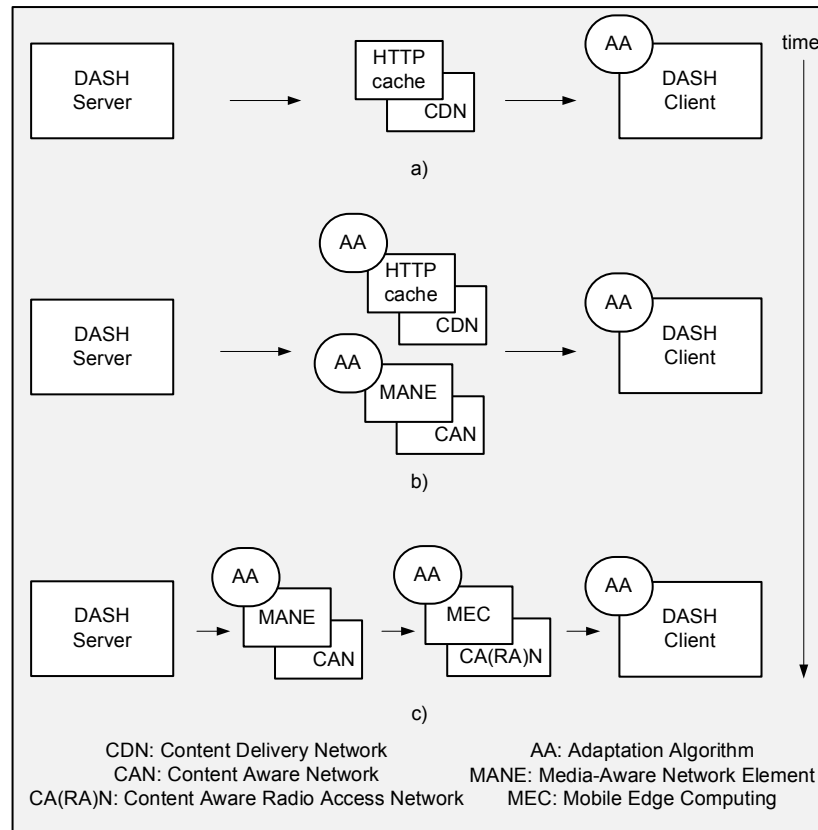


Figure 2. Evolution of media adaptation schemes: a) client-driven adaptations; b) core network-assisted adaptations; c) core and access network-assisted adaptations.

Figure 2a illustrates the traditional DASH adaptation approach based on client decisions. These Over-The-Top (OTT) solutions include end-to-end adaptation loops in order to react to network variability.

Different client-driven adaptation algorithms have been proposed aimed at improving the service performance in different contexts [1, 2, 4]. Furthermore, the advantages and drawbacks of DASH-AVC and DASH-SVC have been largely discussed in the literature [3, 4]. In summary, DASH-AVC results in non-optimal quality selection under fluctuating network conditions, since devices need to a priori choose a unique quality representation. After the decision is made, the whole media segment needs to arrive at the client for its playout. In addition to much faster reaction to variable network conditions, DASH-SVC achieves further quality scalability since already downloaded layers can be enhanced at a later time (e.g., upon quick recovery from degradation).

Initially, the adaptation loop involves the DASH client and the DASH server leading to considerable Round-Trip-Time (RTT) between clients and video sources. Unfortunately, high RTT values degrade the smoothness of adaptation capabilities, with higher impact on DASH-SVC [3].

As a first enhancement, the adaptation loop may end in an HTTP caching node located within the network. As this element is located closer to the end user, RTT in adaptation loops is shortened and the adaptation results in smoother media flows. This scheme also benefits from

Content Delivery Network (CDN) deployments, where the content is served from the closest location to the end user [10].

In summary, DASH-SVC has raised relevant interest in the last years due to several enhancements over DASH-AVC. However, due to the client-driven adaptation approach, none of the adaptation algorithms in the literature are capable of balancing the quality among multiple clients. As discussed in [3], the greedy behavior of DASH clients results in unfair distribution of the available bandwidth and frequent quality fluctuations. This effect could be overcome by implementing cooperative clients or introducing intelligent bandwidth-balancing components inside the network.

Figure 2b represents the next evolution, based on introducing media-aware network intelligence in the DASH provisioning chain [4, 11, 12]. Network-assisted adaptation mechanisms, which cooperate with and/or complement the intrinsic client-driven approach, commence to gain interest in the scientific literature as a way to optimize the media delivery in multi-user scenarios. Special interest has been placed to DASH-SVC, trying to exploit the incremental significance of different video layers and the splitting of layers in different HTTP objects. These proposals address the prioritization of traffic within the core network, either at HTTP caching nodes [4] or at Media-aware Network Elements (MANE) [11, 12].

In this area, MPEG established the Server and Network Assisted DASH (SAND) core experiment in October 2013. SAND standardization is under development as ISO/IEC CD 23009-5 (MPEG DASH Part 5). Through the introduction of DASH Aware Network Elements (DANE), SAND aims at endowing DASH operations with client and network feedback and enabling the addition of adaptation logic beyond the client-driven approach.

Additionally, the Software Defined Networking (SDN) paradigm becomes a relevant candidate to dynamically allocate network resources to different DASH clients in future networks [13]. The SDN manager may be used to allocate content and network bitrates in order to achieve a QoE-based fairness between different users. This approach is well suited for its combination with multi-layer coding in order not to require complex transrating processes.

However, none of these proposals define the interaction with the resource allocation and scheduling mechanisms of underlying radio networks. The design of future 5G networks suggests the standardization of intelligent network nodes that will enable more powerful adaptation and prioritization frameworks over the whole transmission chain, and especially at the edge of mobile network segments [14].

Figure 2c illustrates this trend, aimed at including service adaptation capabilities into the 5G cloud-enabled RAN in order to integrate mobile user awareness and enhance adaptation responsiveness. In this way, specific radio performance parameters could be used to drive the optimization process. In [15], authors propose a 5G SDN manager to allocate on-demand resources to cope with QoE requirements taking into account media characteristics and client buffer status. However, adaptations are triggered by the application-level buffer status and specific radio characteristics are not considered.

To the best of our understanding, all the existing proposals are focused on the resource sharing concerning HTTP caches and wired network elements, or are driven by application-level performance parameters [5]. None of them cope with the specific details of the shared radio access in cellular networks.

This paper presents a novel proposal for optimizing multi-user media delivery within commercial mobile networks. One example of this tendency is the novel MEC industry initiative [9], announced as the ETSI Mobile-Edge Computing Industry Specification Group (ISG) in October 2014, and focused on solving these types of problems from an operator-supported approach. In this way, mobile network operators would provide a standardized open environment where third-party applications may be integrated across multi-vendor platforms. Additionally, MEC will provide an API to third-party partners, offering them access to critical characteristics such as "location awareness" and "network context information". This information may be exploited to develop proximity-enabled services with close-to-zero latency characteristics, in order to optimize the management of future mobile multimedia networks.

Proposed ME-DAF architecture

The proposed novel ME-DAF functional element makes the most out of two convergent trends: the multi-layer enabled Dynamic Adaptive Streaming over HTTP and the Mobile Edge Computing. Figure 3 illustrates the overall network architecture of the resulting solution.

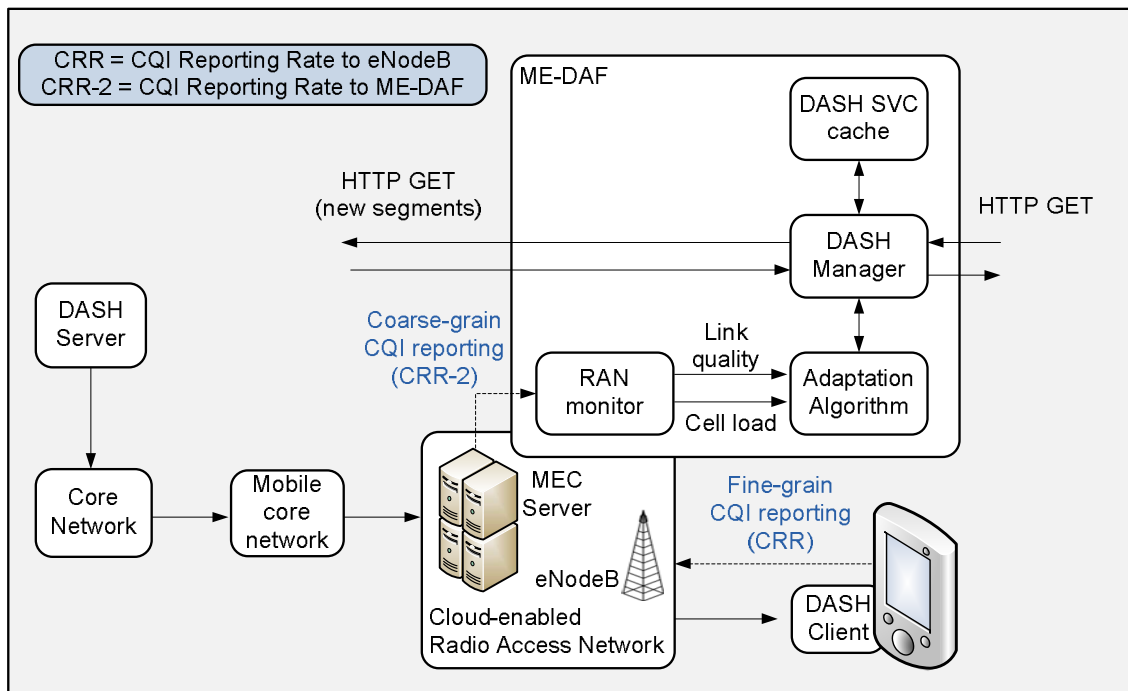


Figure 3. Integration of ME-DAF in C-RAN.

The key element in the MEC industry initiative is the MEC Server, which is integrated into the mobile operators' RAN in order to provide value-added capabilities to third party developers. In the case of LTE, the MEC Server is directly integrated into the eNodeB. The MEC Server is based on cloud computing principles to run third-party applications over a common HW infrastructure. These applications may range from lightweight monitoring instances, which

would provide relevant RAN-related information to optimize external OTT servers, to more complex applications, intended to handle and modify the traffic to/from the mobile users at RAN level.

Upon this MEC Server element, we propose to deploy the ME-DAF element as an intelligent node aimed at optimizing the media delivery to multiple mobile users. ME-DAF is included in the service provisioning chain, thus, enabling the content awareness. At the same time, ME-DAF interacts with the MEC Server in order to retrieve cell and user context information, such as the cell load or per-user radio performance statistics. According to the preliminary technical architecture proposed by the ETSI MEC ISG, the ME-DAF may be deployed as a cloud computing instance within the MEC platform, which is based on C-RAN principles. Therefore, ME-DAF is designed as a RAN-enabled service instance able to capture and manage the HTTP transactions of the associated customers. Depending on the business model, ME-DAF could be deployed to operate for a specific OTT operator or for a mobile network operator (including virtual operators).

The proposed ME-DAF is aimed at orchestrating the delivery of HTTP objects to the different multimedia users present in the cell. According to the state of the art, multi-layer enabled client devices would select the best achievable media representation at each time. In case of sudden individual radio degradations or cell saturation, an element such as ME-DAF would be able to cope with the dynamic adaptation of sessions. As a result, different media segment requests arrive at ME-DAF, which needs to prioritize the delivery of HTTP objects to maximize the overall performance. Consequently, ME-DAF adds multi-user service logic over the traditional standalone client-driven adaptation approach, being able to prioritize and/or drop different HTTP transactions according to a QoE-driven fair scheduling of radio resources.

For this aim, ME-DAF takes into consideration the significance of each media segment (base layer and enhancement layers), the current radio conditions of each user (and thus, the actual achievable quality), and the available resources in the cell for the service.

Figure 3 illustrates the main functional modules of ME-DAF:

- DASH Manager is in charge of handling media segment requests from the different clients, triggering HTTP caching functions in order to increase both the network efficiency in the wired elements and the user-experienced latency. In this sense, DASH Manager is responsible of replying to HTTP GET requests made by the mobile users, and forwarding these requests to the external media source when the content is not previously cached. The DASH Manager also analyzes the SVC-enabled MPD files and instructs the Adaptation Algorithm module with the flow-level characteristics of the traffic. Finally, it implements the decisions of the Adaptation Algorithm and performs the media delivery towards the end users.
- DASH SVC cache module is in charge of storing and serving media segments as instructed by DASH Manager.
- The Adaptation Algorithm is in charge of making QoE-driven scheduling decisions in order to maximize the service experience. This module combines content awareness, user context awareness and cell status awareness. In this way, the ME-DAF is able to exploit the benefits of the multi-layer media delivery to incorporate network-assisted

adaptations to the client-driven approach. In order to perform the channel aware estimations, the Adaptation Algorithm is fed with low-level detailed information provided by the MEC Server. Being the most radio technology-dependent part of ME-DAF, the Adaptation Algorithm module needs to be adapted to each specific deployment.

- The RAN Monitor module is in charge of implementing the MEC API to interact with the MEC Server in order to retrieve cell statistics and individual channel state information (CSI), which is provided by means of Channel Quality Indicator (CQI) in LTE. CSI is reported by UEs to eNodeB at CQI Reporting Rate (CRR), usually configured to low values (e.g., 5 ms). In multi-user scenarios, this feedback granularity would entail a high traffic volume in the interface between MEC Server and ME-DAF RAN Monitor. Thus, a coarse-grain feedback statistic is defined in Fig. 3 as CRR-2.

According to the proposed architecture, RAN Monitor is the only module mandatorily deployed at C-RAN. This element could retrieve the low-level information and forward it to another ME-DAF instance running at any of the wired segments. This flexible solution provides the required scalability to adapt ME-DAF to the available virtualized computing resources. The proposed architecture allows partially deploying ME-DAF outside the C-RAN, which would entail economic savings at the cost of increased RTT and reduced media adaptation responsiveness.

Performance evaluation

In this section, we analyze the performance of the proposed ME-DAF element under DASH-SVC transmissions in modern cellular networks. For that purpose, we compare through simulations the achieved performance in terms of QoE of the following approaches:

- eNodeB scheduling strategy, without further higher-layer adaptations, aimed at analyzing the application-level impact of LTE schedulers in scenarios with shared radio resources. Two types of eNodeB scheduling functions are evaluated:
 - Best CQI (BC). The scheduling function prioritizes those UEs with better radio conditions, i.e. higher reported CQI values. Thus, the objective is to maximize the overall achieved throughput of the cell at the cost of reduced fairness between UEs.
 - Layer CQI (LC). The scheduling function schedules the transmission of base layers first and the transmission of enhancement layers later [11]. Within each priority class, the scheduler applies BC principles among UEs. This type of eNodeB aims at maximizing the number of UEs with a minimum level of quality.
- Combined client-driven adaptation and eNodeB scheduling, aimed at gauging the improvement of DASH clients in the considered network conditions. The clients implement typical buffer-aware adaptation capabilities, forcing them to decrease the requested video quality upon foreseeing buffer starvation situations.
- Complex ME-DAF based adaptations, jointly to client-driven adaptations over the considered eNodeB scheduling strategies, aimed at evaluating the enhancements of

the proposed architecture. Two types of radio feedback parameters are evaluated, leading to two ME-DAF implementations:

- CQI-aware MEC (C-MEC): ME-DAF receives periodic CQI values for every involved UE.
- CQI- and Block Error Rate (BLER)-aware MEC (CB-MEC). In addition to CQI, the MEC Server is able to trigger individual information when UEs experience high BLER due to high mobility patterns.

For performance evaluation, we simulate the behavior of an LTE network in MATLAB. The cell is configured with a bandwidth of 20 MHz, serving a cluster of Category 4 devices (up to 150 Mbps in downlink with MIMO). Since this paper aims at evaluating how ME-DAF enhances other strategies in real-world deployments, even with inaccurate CSI information, we develop a trace-based simulator instead of running complex physical LTE models. Thus, we simulate the performance of the radio channels by including experimentally gathered CQI traces, as shown in Fig. 4.

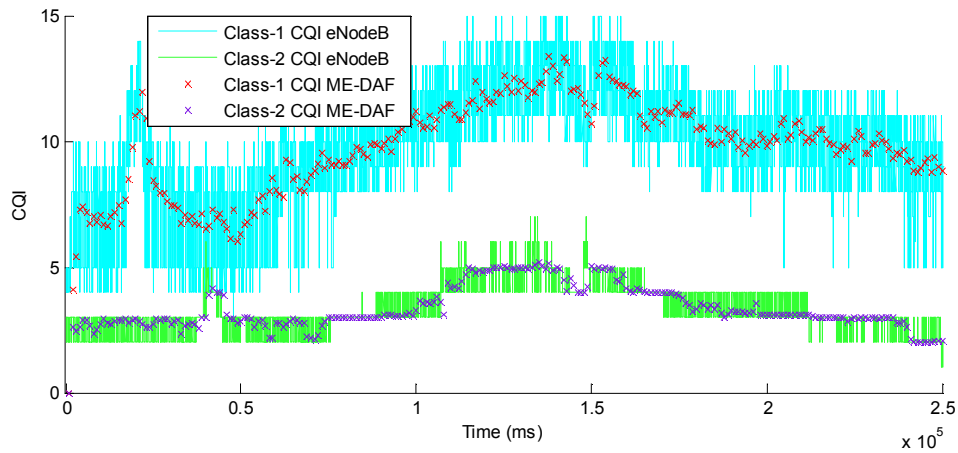


Figure 4. CQI traces at eNodeB and ME-DAF.

Two classes of UEs are considered, resembling better (Class-1 UEs) and worse (Class-2 UEs) channel conditions. Experimental CQI traces at eNodeB provide a granularity of 5 ms, while CQI traces available at ME-DAF are generated from averaging these traces with 1 s granularity. Average CQI values are 9.47 and 3.36 for Class-1 and Class-2 UEs respectively, resulting in average data rates of 67,326.56 kbps and 13,651.2 kbps.

The implemented LTE schedulers (BC and LC) arrange the transmissions of multi-user traffic at every Transmission Time Interval (TTI) of 1ms, allocating the set of available Resource Blocks (RBs) according to traffic demands and CSI reports. Based on the reported CQI values, eNodeB applies Adaptive Modulation and Coding (AMC) to select the most suitable Modulation and Coding Scheme (MCS) for each UE according to 3GPP TS 36.213. Additionally, eNodeB applies conservative MCS assignments when high BLER is detected in order to mitigate the effect of radio retransmissions under highly variable radio channel conditions.

Regarding traffic demands, we consider a fixed number of UEs in the cell accessing different SVC video sessions. Each video session is composed of series of Group of Pictures (GOPs), and each GOP comprises a base video layer (L0) and two enhancement video layers (L1 and L2). We

configure the GOP length to 2s and flow size distributions are characterized from traces provided in [1]. The mean size of base layer, first enhancement layer and second enhancement layer are 73.97 KB, 161.63 KB and 422.66 KB respectively. Taking into account the GOP size, the average video bitrates for each aggregated layer are 295.9 kbps, 942.4 kbps and 2,633.04 kbps. Audio-related information is included as another media representation, considering exponentially distributed audio segment sizes with mean size of 25 KB (for an average of 100 kbps). Therefore, the whole video stream would require an average of 2,733.04 kbps for its complete transmission.

From a QoE perspective, a media segment should be downloaded on time for its smooth playout. The trade-off between responsiveness and smoothness is driven by the client buffer size, which establishes the maximum transmission delay tolerable to avoid video stalling and re-buffering effects [5]. In our case, the buffer size is configured to 10 s and video segments with higher transmission delays are discarded. The playable video representation is then mapped to QoE following SSIM values provided in [1] and Mean Opinion Score (MOS) mappings proposed in [7]. Although being a subjective metric in nature, MOS is typically also used as a metric to estimate the impact of different objectively measurable aspects in the QoE. In [5], several references are analyzed to show how MOS can be properly mapped to initial playout delay and video stalling patterns, image quality encoding levels, image spatial resolution and image temporal resolution. In summary, we configure MOS levels achievable for each video representation as: 3 for L0, 3.5 for L1 and 5 for L2. In case only audio or no media segment is received, we assign a MOS value of 2 and 1 respectively.

Finally, the adaptation algorithm implemented for performance evaluation falls in the field of traditional multi-objective optimization. The problem becomes a single-objective utility function considering averaged MOS scores for the set of users, constrained by the minimum acceptable MOS level of 3. The algorithm takes as inputs the overall available RBs and CQI values for each user, while the set of allowed content bitrates is a system parameter. The optimization vector is composed by the number of RBs for each user, constrained by the possible content bitrates and the current MCS. The adaptation algorithm is implemented as a genetic algorithm in MATLAB and runs at GOP scale with sufficient converging time.

Achieved QoE levels

Taking into account both CQI traces and video patterns, we show simulation results with 20 mobile users equally assigned to Class-1 and Class-2 UEs. In this scenario, the considered LTE cell is forced to congestion and the different media adaptation algorithms can be evaluated in severe load conditions.

Figure 5 shows the ratio of correctly received video segments in function of different media representations (i.e., audio, video base layer and video enhancement layers). Each subplot includes all the information concerning the evaluated approaches. “eNodeB” denotes the case where only eNodeB scheduling influences the transmission of media flows. “Client” denotes the case where UEs implement client-driven adaptations. Finally, “C-MEC” and “CB-MEC” represent the two proposed ME-DAF approaches. For each scheduling policy (BC and LC), results for Class-1 UEs, Class-2 UEs and aggregated values are offered.

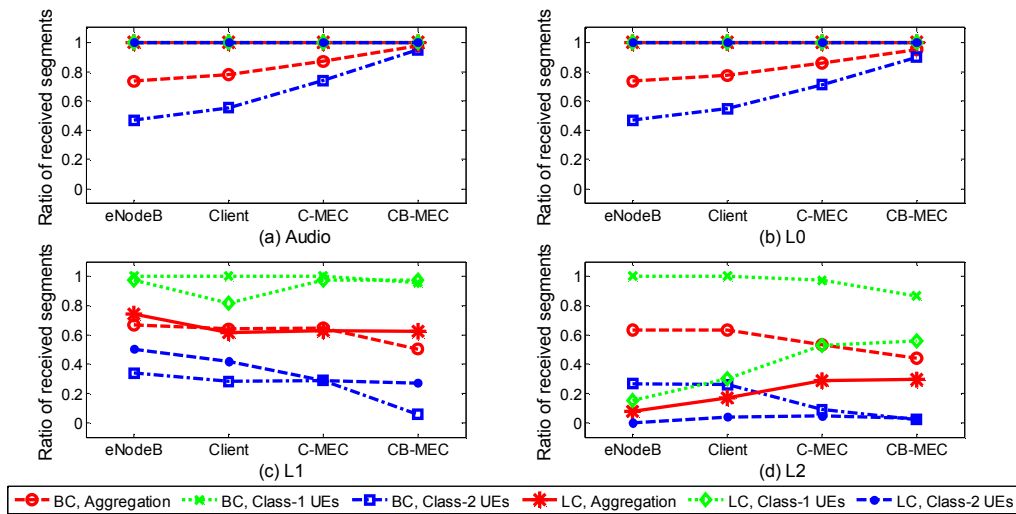


Figure 5. Ratio of correctly received media segments: a) audio; b) base layer (L0); c) enhancement layer 1 (L1); d) enhancement layer 2 (L2).

As can be observed, BC-driven eNodeB transmissions result on clear unfairness between UE classes. While Class-1 UEs receive 100% of media segments at highest quality level, Class-2 UEs do not reach a sufficient quality level since they receive less than 50% of media segments belonging to audio and video base layer. Thus, one-half of mobile customers would be clearly unsatisfied with the service.

LC-driven eNodeB provides the inverse trend. In this case, eNodeB tries to transmit higher priority media segments (i.e., audio and video base layer) until they are discarded due to application-level playout threshold. Therefore, 100% of audio and base layer media segments properly arrive at Class-1 and Class-2 clients. Consequently, eNodeB tries to deliver L1 for Class-2 UEs before L2 for Class-1 UEs, which leads to inefficient use of radio resources. In this case, only 20% of L2 video segments are delivered to Class-1 UEs.

Client-driven buffer-aware adaptation enhances the quality of media sessions to some extent, through a better use of the assigned radio resources. Under BC-driven eNodeB conditions, only Class-2 UEs need to dynamically adapt the requested media quality. As a result, Class-2 UEs avoid requesting L1 media segments to allow the transmission of L0 segments of consecutive GOPs. Under LC-driven eNodeB, both Class-1 and Class-2 UEs need to apply quality adaptations and the behavior is similar to the previous case.

In order to better understand the effect of these results on QoE, Fig. 6 illustrates the mapping of the playable quality levels into MOS scale. Additionally, MOS allows a better understanding of the behavior and effect of ME-DAF.

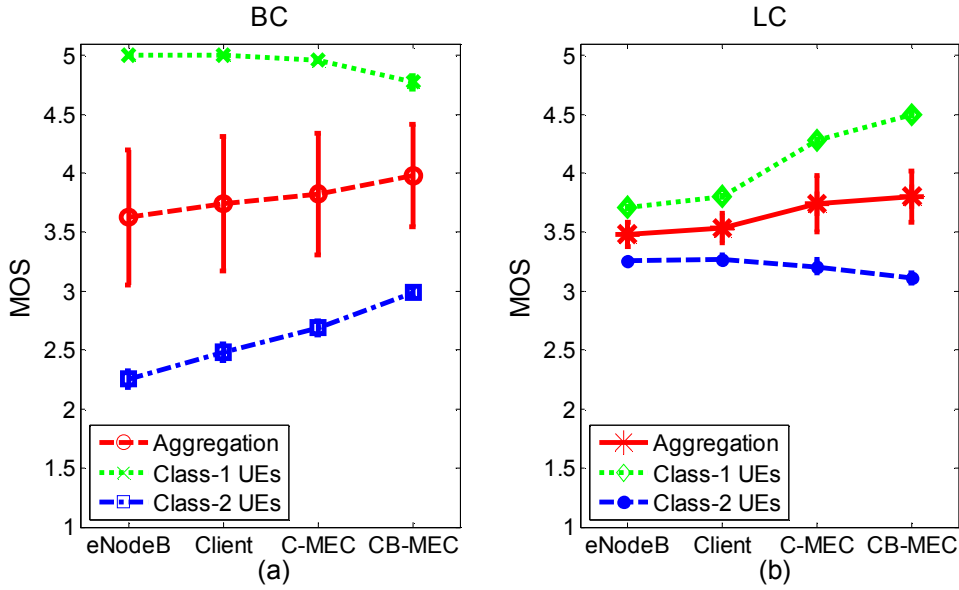


Figure 6. Resulting MOS: a) eNodeB Best CQI; b) eNodeB Layer CQI.

The proposed scenario illustrates the limitations of individual decision making in client-driven adaptations, and the lack of QoE awareness in eNodeB scheduling policies.

The proposed ME-DAF scheme approximates the optimal solution in terms of QoE. Under BC-driven eNodeB, ME-DAF is capable of dropping L2 media segments for Class-1 UEs in benefit of Class-2 UEs, which are able to reach the established lower MOS threshold. Under LC-driven eNodeB, ME-DAF is capable of dropping L1 media segments for Class-2 UEs in benefit of Class-1 UEs, which experience an increase on the number of received L2 media segments.

As a result, we observe how the optimal solution is approximated by different fronts, showing the efficient performance of the proposed ME-DAF adaptation approach in different LTE deployment scenarios.

Conclusions

This paper proposes a novel architecture for improving the efficiency in the transmission of DASH traffic over modern mobile networks. We introduce a new element, denoted as ME-DAF, which is placed within the C-RAN to add enhanced network support to the multimedia service. The associated delivery scheme entails both client-driven and mobile network-assisted adaptation of media sessions, providing a quick and smooth reaction to typical performance variability and resource sharing problems in multi-user mobile networks.

Our proposal benefits from state of the art technologies including DASH-SVC and MEC. The former allows quicker reactions to sudden performance degradations while maintaining a higher average media bitrate. Additionally, it enables a more efficient network-assisted adaptation based on prioritized scheduling strategies. The latter allows an improved interaction with operators' RAN, providing access to radio performance statistics and enabling the deployment of mobile services close to end users.

ME-DAF implements content awareness, in the sense that it captures the significance of the different media flows, user awareness, since it takes into account the current radio conditions of the different users, and network awareness, through monitoring the available resources in the cell.

We evaluate the performance of the proposed ME-DAF element against state of the art eNodeB and client-driven adaptations. ME-DAF is capable of conforming media sessions to maximize the average QoE, taking as constraint the minimum quality threshold. As a result, we may state that the proposed multi-user adaptation algorithm is well-suited to exploit the awareness of dynamic radio conditions in multi-user mobile multimedia scenarios.

Research challenges

The presented results illustrate the applicability of the MEC principles for an efficient multi-user DASH-SVC media delivery over currently deployed 4G network architectures. These promising results open a series of research challenges to be covered over the next years:

- The concept of QoE may be generalized to include more performance parameters [5], other content types and formats, and user contextual factors [6]. The proposed architecture is valid to incorporate evolved adaptation algorithms coping with extended functionalities.
- The interactions between client and network adaptations need to be further studied. Although some recent studies address the problem, it is not clear how the different control loops may impact the QoE.
- The current adaptation algorithm takes into account LTE CSI in terms of CQI values, and implements QoE-aware fair resource sharing with the knowledge of currently deployed LTE MIMO technologies. The evolution of 4G will bring new radio technologies, including enhanced modulation and beamforming, multi-user MIMO and Coordinated MultiPoint (CoMP). ME-DAF needs to be aware of the specific technology used for each mobile device and the specific CSI parameters, in order to implement optimal decisions. Additionally, in the case of CoMP and other heterogeneous networks, ME-DAF would need to orchestrate its behavior over different eNodeB's following a centralized RAN approach.
- Future 5G networks will also bring new challenges, including radio technologies such as millimeter wave (mmWave) technologies, new small cell heterogeneous architectures and the use of Network Function Virtualization (NFV). The design of an evolved ME-DAF requires previous standardization of the 5G architecture (including CSI) and the integration of MEC into 5G.
- The integration of this kind of MEC-based elements with future MPEG DASH and 3GPP standards is an open issue.
- ME-DAF supports different business models, in terms of ownership, user segmentation and collaboration between OTT and network operators. Specific studies are needed to evaluate its performance in different scenarios devised by the ETSI MEC ISG.

References

- [1] C. Sieber *et al.*, "Implementation and User-centric Comparison of a Novel Adaptation Logic for DASH with SVC," *Proc. IFIP/IEEE International Symposium on Integrated Network Management*, 2013, pp. 1318–1323.
- [2] C. Muller *et al.*, "Using Scalable Video Coding for Dynamic Adaptive Streaming over HTTP in mobile environments," *Proc. IEEE 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 2208–2212.
- [3] J. Famaey *et al.*, "On the merits of SVC-based HTTP adaptive streaming," *Proc. IFIP/IEEE International Symposium on Integrated Network Management*, 2013, pp. 419–426.
- [4] R. Huyssegems *et al.*, "SVC-Based HTTP Adaptive Streaming," *Bell Labs Tech J*, vol. 16, no. 4, 2012, pp. 25–41.
- [5] M. Seufert *et al.*, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Commun Surv Tut*, vol. 17, no. 1, 2015, pp. 469–492.
- [6] K. Mitra *et al.*, "Context-Aware QoE Modelling, Measurement, and Prediction in Mobile Computing Systems," *IEEE T Mobile Comput*, vol. 14, no. 5, 2015, pp. 920–936.
- [7] T. Zinner *et al.*, "Towards QoE management for scalable video streaming," *Proc. 21th ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE*, 2010, pp. 64–69.
- [8] J. Wu *et al.*, "Cloud Radio Access Network (C-RAN): A Primer," *IEEE Network*, vol. 29, no. 1, 2015, pp. 35–41.
- [9] M. Patel *et al.*, "Mobile-Edge Computing Introductory Technical White Paper," 2014.
- [10] S. Lederer *et al.*, "Adaptive Multimedia Streaming in Information-Centric Networks," *IEEE Network*, vol. 28, no. 6, 2014, pp. 91–96.
- [11] N. Bouten *et al.*, "Deadline-based Approach for Improving Delivery of SVC-based HTTP Adaptive Streaming Content," *Proc. IEEE Network Operations and Management Symposium (NOMS)*, 2014, pp. 1–7.
- [12] M. Grafl *et al.*, "Scalable Media Coding Enabling Content-Aware Networking," *IEEE MultiMedia*, vol. 20, no. 2, 2013, pp. 30–41.
- [13] P. Georgopoulos *et al.*, "Towards Network-wide QoE Fairness Using OpenFlow-assisted Adaptive Video Streaming," in *Proc. ACM SIGCOMM workshop on Future human-centric multimedia networking (FhMN '13)*, 2013, pp. 15–20.
- [14] D. Soldani and A. Manzalini, "Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society," *IEEE Veh Technol Mag*, vol. 10, no. 1, 2015, pp. 32–42.
- [15] C.-F. Lai *et al.*, "A Buffer-Aware HTTP Live Streaming Approach for SDN-Enabled 5G Wireless Networks," *IEEE Network*, vol. 29, no. 1, 2015, pp. 49–55.

Biographies

JOSE OSCAR FAJARDO (joseoscar.fajardo@ehu.es) works as research fellow in the Department of Communications Engineering of the University of the Basque Country (UPV/EHU), at the Faculty of Engineering in Bilbao. He received his M.Sc degree in 2003. He works in adaptive management of mobile multimedia services under the framework of IMS. He has co-authored more than 25 journal and conference papers since 2005, mainly in areas of QoS/PQoS/QoE and service performance assessment, and QoS-aware networking.

IANIRE TABOADA (ianire.taboada@ehu.es) works as lecturer in University of the Basque Country (UPV/EHU). She received her degree and her Ph.D. in Telecommunications Engineering from UPV/EHU, in 2008 and 2013 respectively. Her research interests mainly include dynamic scheduling of network resources, especially for QoE optimization in wireless

networks. She is familiar with Markov Decision Processes, Gittins index approach and Marginal Productivity indices (Whittle index theory).

FIDEL LIBERAL (fidel.liberal@ehu.es) received his BS and MS in Telecommunications Engineering from the University of the Basque Country (UPV/EHU), Spain, in 2001. In 2005, he received his PhD in Telecommunications Engineering from UPV/EHU in the area of holistic management of quality (both QoS and QoE) in telecommunications services. He currently works as a Lecturer and researcher in the Faculty of Engineering of Bilbao, where he has co-authored more than 35 conference and journal papers.