Enabling Layered Video Coding for IMS-Based IPTV Home Services

Ivan Vidal, Jaime Garcia-Reinoso, and Francisco Valera, University Carlos III of Madrid
Alex Bikfalvi, IMDEA Networks

Abstract

Nowadays IPTV services are gaining attention from both providers and end users. There is a large effort toward the integration of these services into emerging next-generation network architectures. In particular, one of the most relevant solutions is being proposed by ETSI-TISPAN and is based on the IP multimedia subsystem. This article focuses on introducing layered video coding into TISPAN IMS-based IPTV architecture, allowing cost-effective efficient solutions both for residential users and providers (e.g., flexible support of heterogeneous devices, live mosaics, adaptive video quality based on device and/or network capabilities). The advantages of using layered video coding in the TISPAN IPTV solution are analyzed and illustrated with a set of use cases. Furthermore, this solution has been integrated into a multimedia testbed in order to validate the presented proposal.

IPTV services are key business enablers for service and access providers, but it is very difficult to deploy these services in the current Internet (further than the video distribution most asymmetric digital subscriber line [ADSL] providers are now offering) due to diverse problems related to the lack of quality guarantees (best effort Internet). These problems are partially solved by some network architectures defining a complete architecture for several services including IPTV. In this respect, the European Telecommunications Standards Institute Telecommunications and Internet Services and Protocols for Advanced Networks (ETSI-TISPAN) work group has an important role, developing a set of specifications for a next-generation network (NGN) architecture. One of the solutions provided by TISPAN is to use an IP multimedia subsystem (IMS) core to provide IPTV services (i.e., IMS-based IPTV). Although this may have implications on the transport provider network more than on the residential framework, one of the most interesting issues for this article about IMS-based IPTV is that it uses several IP multicast trees, one per TV channel, to reach user devices. If the service provider offers several video qualities for a mobile phone, a laptop, and a television, each quality must use a different multicast tree, and each device should be subscribed to a different multicast group. This technique is not new, and is called *simulcast*: the terminal, depending on its capabilities, chooses one multicast tree providing the desired channel/quality. The most important characteristic in IMS-based IPTV (and in the TISPAN NGN in general) compared to the best effort Internet model is that sessions are established using Session Initiation Protocol (SIP) and resources are reserved in network elements between servers and users, guaranteeing the quality of experience (QoE) for the user.

In this article we propose to enhance the IMS-TV service by supporting *layered video multicast*. This technique allows splitting a single video into different streams (layers), and its inclusion into the IPTV TISPAN architecture has important advantages that can be illustrated with the following use cases.

Heterogeneous home environment: In an NGN environment, a possible IPTV home scenario may involve heterogeneous clients connected to the same video server (set-top box, mobile phone, laptop, etc.). When using layered coding, the service provider will have to consider the number of layers only for a specific channel, not for each specific device (a mobile phone with limited resources may just be able to receive a single layer, while a television may be able to receive more and provide better quality).

Video quality update: If the user is watching a TV channel with a certain quality and wants to change the quality (or the device automatically adjusts it), the service can be interrupted in case of simulcast because it implies leaving the current multicast tree and joining the multicast tree associated with the new quality. Using layered coding, this is easy: to increase the quality, it is enough to keep on receiving the current layers, and join additional layers up to the desired quality; to decrease the quality, the process would consist of leaving as many multicast trees as necessary to achieve the targeted quality.

Reduced bandwidth: A user is watching a TV program on a high-definition TV and wants to record it on a home personal video recorder (PVR) so her son can watch it on his mobile phone later. Nowadays, with the TISPAN proposal, both an IMS television set and a home PVR would have to connect to two different multicast trees for the same channel (e.g., high definition and standard definition). Depending on the user downlink bandwidth, this may be impossible. Had we used layered coding, only the layers for the high-definition video would be required, and a subset of them would be used by the PVR.

Effective channel change: In TISPAN the IMS core reserves resources for the maximum possible bit rate, despite of the fact that the user is just watching a standard-definition channel (they prefer to do bandwidth overbooking in order to

 0890-8044/09/$25.00

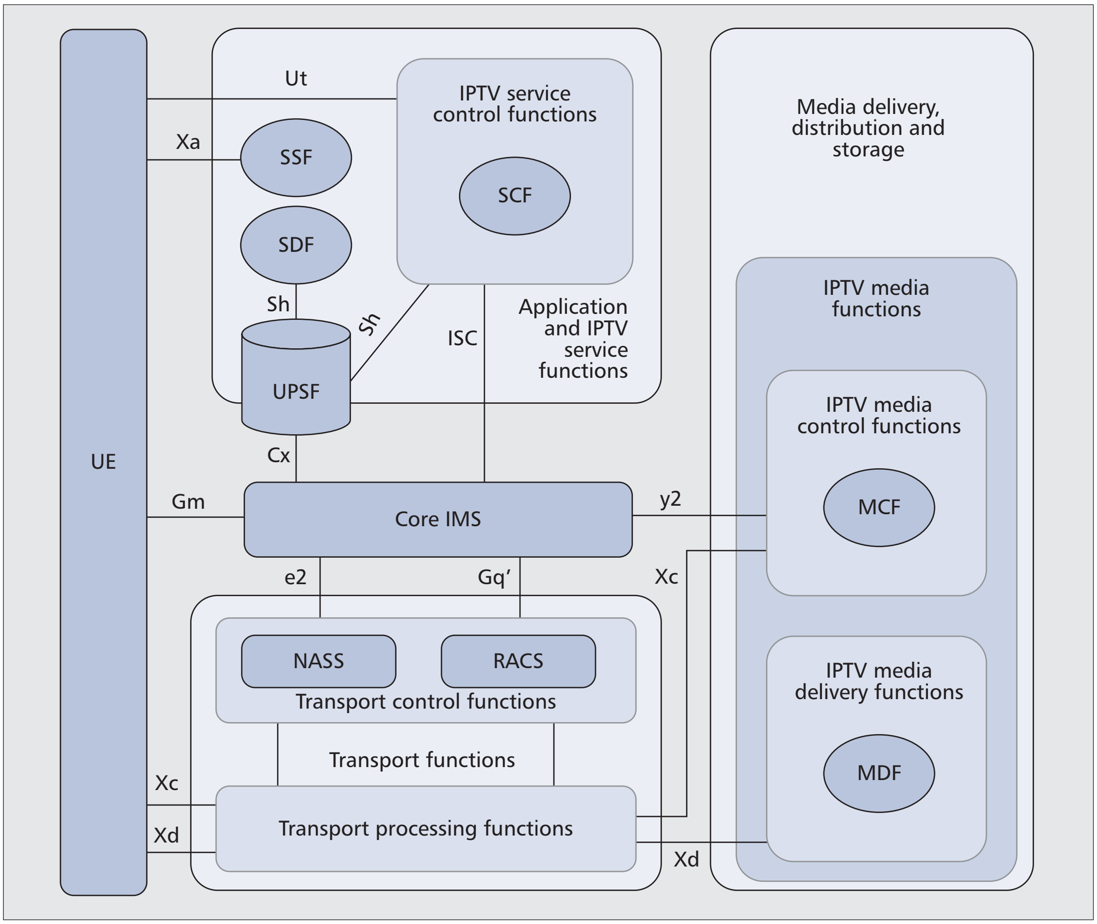

Figure 1. *Functional architecture for IPTV services.*

obtain a better channel change delay). With layered video coding, the system will only have to reserve the necessary bandwidth for the tuned channel, preserving the channel change delay provided by the TISPAN mechanism.

Enhanced live mosaics: Live mosaics with low bandwidth, allows presenting several video channels (not just pictures) and fast switching after selecting a channel. To display the mosaic, the IMS TV device joins one layer for each channel. After the user chooses a particular channel, the IMS-TV leaves the multicast trees for the other channels in the mosaic, and only joins as many layers as necessary for the selected channel. This can be done with the TISPAN proposal, but with layered coding we can effectively increase the number of channels in the mosaic, with the same user equipment (UE) capabilities and available network resources.

The article is organized as follows. The next section introduces the IPTV TISPAN architecture. We then present an overview of layered video coding. The following section covers a detailed description of all the steps required to introduce layered video coding in the IMS-based IPTV architecture and an analysis of the advantages over the TISPAN proposal. We then present a basic testbed developed using our proposal, while the final section concludes highlighting the main contributions of this article.

TISPAN IMS-Based IPTV

Figure 1 shows the functional architecture defined by TISPAN for IMS-based IPTV services [1]. This architecture has been designed to accommodate the diverse IPTV services being considered in TISPAN specifications: content on demand (CoD), broadcast (BC) TV, and network PVR (N-PVR).

The architecture is composed by a set of functional entities and high-level functions: the UE, *core IMS*, *transport functions*, *user profile server function* (UPSF), *application and IPTV service functions*, and *media delivery, distribution, and storage functions*. These functional entities are interconnected by a set of related interfaces (further details can be found in [1]).

UE: An IPTV enabled NGN terminal capable of handling the control and media flows related with the IPTV service. The UE implements the interaction with the end user, providing the user with the means to browse and select among the available services (TV channels, CoD listings, etc.).

Core IMS: A subsystem located at the service layer in the TISPAN NGN architecture that supports the provision of SIP-based multimedia services to the UE. The core IMS is a subset of the IMS, as defined by the Third Generation Partnership Project (3GPP) [2], restricted to the session control functionalities.

Transport functions: Provide IP connectivity to NGN terminals. These functions can be further decomposed into a set of *transport control functions* and a set of *transport processing functions*.

Within the transport control functions, the resource and admission control subsystem (RACS) acquires a special relevance. It performs policy control, resource reservation, and admission control functions for *unicast* and *multicast* traffic for transport networks in the NGN. The RACS provides applications with a means to reserve transport resources and guarantee QoS for the value-added services in the NGN.

UPSF: The UPSF hosts the IMS user profile. In addition, it can also contain an IPTV user profile. This profile covers the user settings necessary to operate the IPTV service, such as language preferences or the list of subscribed TV channels.

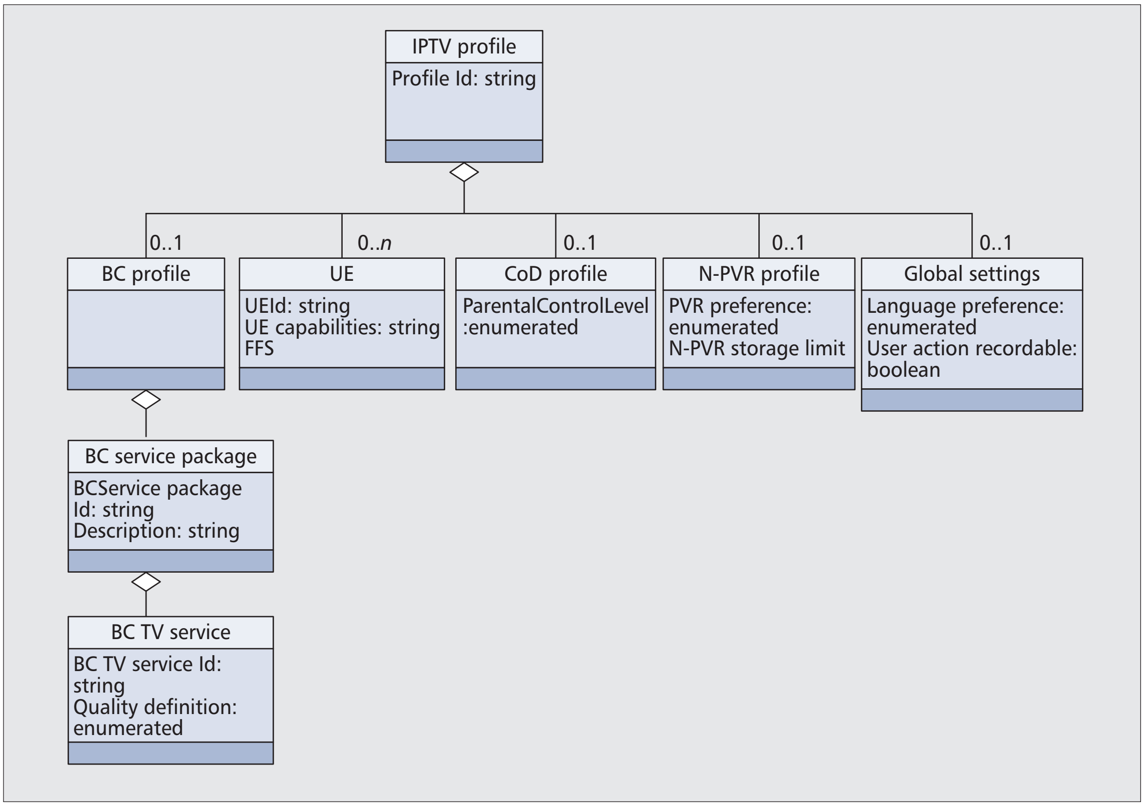

Figure 2. *IPTV user profile.*

Application and IPTV service functions: This set of functions enables the operation of IPTV service. It includes the following functional entities: the *service discovery function* (SDF), *service selection function* (SSF), and IPTV *service control function* (SCF).

The SDF provides service attachment information to the UE, containing SSF addresses. The communication between the SDF and the core IMS follows the specifications for the ISC reference point, and is based on SIP. Two modes of operation have been defined to provide the UE with service attachment information: a push mode, where the SDF actively sends the information to the UE when it attaches to the network, and a pull mode, where the UE actively requests this information from the SDF after attaching to the network. Finally, the SDF can personalize the service attachment information that is provided to the UE, for instance, by using the UE location information, the user subscription information within the IPTV service, and the UE capabilities (vendor, model, etc.).

The SSF provides the UE with service selection information, such as a list of TV channels within the BC service that the UE can browse and select. The communication between the SSF and the UE is based on HTTP.

The UE can indicate personalization information to the SSF when requesting service selection information. With this information, the SSF can fetch the IPTV user profile, and customize the service information that is returned to the UE. For instance, a parental control level (if included in the user profile) could be used to delete inappropriate elements from the returned CoD listing, or the capabilities associated with the requesting UE (e.g., supported encodings and frame rates) could be used to return the information of those TV channels that can be properly played by the UE.

The SCF is a SIP application server (AS) that implements the IPTV service logic. The SCF is in charge of authorizing the user access to the IPTV service, and selecting, if necessary, the appropriate *IPTV media functions*. In order to access the IPTV service, the UE needs to execute certain SIP-based session control procedures that involve the SCF. The communication between the UE and the SCF to accomplish these procedures takes place via the core IMS over the ISC reference point. In addition, the UE may use the Ut reference point, based on HTTP, to manage the IPTV user profile (this profile is also kept at this location).

Media delivery, distribution, and storage: Related to this set of functions, two functional entities acquire significant relevance: the *media control functions* (MCF) and *media delivery functions* (MDF). The MCF mainly provide MDF management functionalities and handles the interaction with the SCF, while the MDF essentially implement media handling functionalities, such as storing, processing, and delivery.

The IPTV User Profile

Figure 2 describes the structure of the IPTV user profile, as defined in [1].

The profile contains a set of object classes specific to the CoD, BC, and N-PVR services. Focusing on the BC service, the IPTV user profile can contain a BC profile, comprising one or more BC service package descriptions. A BC service package is a set of BC TV services related to a particular TV channel and an associated quality definition. The BC profile only contains a reference to the BC service packages and BC TV services to which the user has subscribed.

Accessing the BC Service

In order to access the IPTV service and start receiving the media streams corresponding to BC TV service, a session initiation procedure must take place between the UE and the SCF. Figure 3 shows an example of the SIP signaling flow corresponding to the session establishment, assuming that the session is initiated from the UE. Detailed procedures for the involved protocols are described in [3]. It is assumed that the

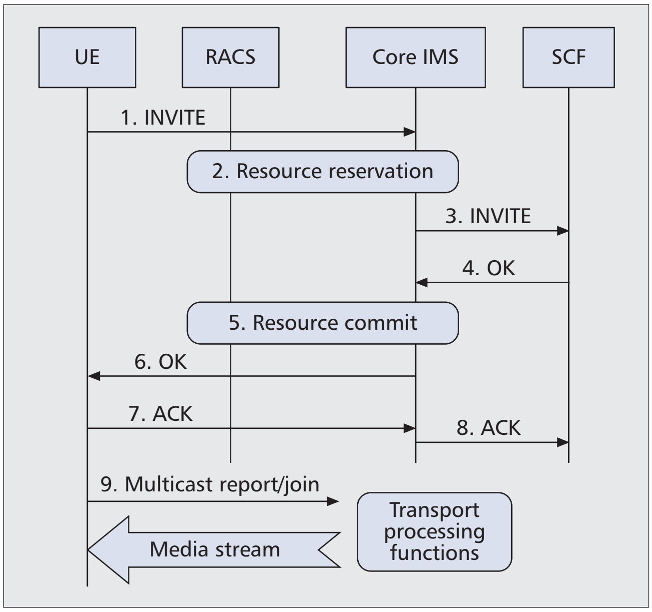

Figure 3. *Session initiation for BC service.*

UE has previously retrieved from the SSF the service selection information necessary to access the BC service.

The UE generates a `SIP INVITE` request containing an SDP offer (step 1). This SDP offer will include a media description, where the UE indicates, among other things, the BC service the UE wants to join first, the multicast address corresponding to the BC service, the bandwidth requirements for the session, and the set of service packages to be authorized during the session. It is important to note that the bandwidth requirements are defined by the largest bandwidth of all the BC services included in the session.

The `INVITE` request is sent toward the core IMS, which contacts the RACS in order to make a proper resource reservation for the session (step 2). Eventually, the request is received at the SCF. Then the SCF checks the requested service packages against the subscription information contained in the IPTV user profile. At this point, the SCF can restrict the service packages and BC services that can be accessed during the session. Finally, the SCF answers the request with a `SIP OK` response (step 4), which includes an SDP answer with the outcome of the authorization process. The `OK` response is sent to the core IMS, which contacts the RACS to modify the resource reservation (if necessary), commit this reservation, and activate the service packages in the transport network, thus enabling multicast joining. Finally, the `SIP OK` response is sent to the UE, which generates and sends a `SIP ACK` request toward the SCF (step 7). At this point, the UE can join the multicast TV channel (step 8).

For channel changing, the UE simply leaves the multicast TV channel and joins the new channel, by means of the Internet Group Multicast Protocol (IGMP). The SCF does not participate in this procedure, so as to avoid large zapping delays.

Layered Coding Overview

The concept of layering coding is not new, and there are many initiatives trying to define better codecs to increase overall performance. With layered coding, the codec decomposes the original video in several non-overlapping streams or layers, using one of the two following layering approaches: *cumulative* (or scalable video coding [SVC]) [4, 5] and *non-cumulative* (or multidescription coding [MDC]) [6]. In the cumulative schemes, there is one layer called the *base layer* that is necessary in order to decode others. In other words, if the base layer is not received, the remaining layers are not useful to decode the video. On the other hand, non-cumulative techniques generate layers of the same relevance, and no matter which layer is received, it contributes to increase the quality perceived by the user.

Some studies [7] have confirmed that three to five layers are enough to have reasonable fairness. This means it is necessary to have at least three layers to completely define the original video with a reasonable number of descriptors, but more than five layers is not recommended because the overhead introduced by the codification is not justified.

The overhead is another interesting issue for both SVC and MDC. As both methods introduce redundancy in all generated layers, the resultant video rate is bigger than the single-layer video version. This overhead goes from low values, 23 percent for SVC, to big ones, 44 percent for MDC [7].

As can be inferred from these results, it is always better to transmit single-layer videos in terms of bandwidth, but depending on the scenario, it could be useful to use a layered video coding technique. When all possible paths are equal in terms of losses or delay, MDC is preferred (typical for wireless mesh or even peer-to-peer applications for streaming), but if it is possible to have one high- priority channel, SVC is recommended.

Layered Multicast in TISPAN IMS-Based TV

This section describes a set of proposals to enable layered multicast in the BC service architecture proposed by TISPAN. As a design goal, these changes must not impose any architectural modifications in the TISPAN IMS-based TV specifications.

With layered coding, the video corresponding to a single TV channel is decomposed into a set of substreams that will be delivered to the UE by means of network multicast. This way, assuming that the user has subscribed to a given BC TV service with a certain quality, instead of transmitting the service by means of a single multicast tree, the service will be split into a set of non-overlapping substreams associated with different multicast trees. These substreams will jointly provide the service with the quality definition level specified in the IPTV user profile.

Therefore, assuming that the BC TV services make use of layered coding, the service selection information provided by the SSF to the UE must include the new parameters that allow the UE to receive the different substreams corresponding to a subscribed BC TV service. In this respect, the XML schemas defined in [8] need to be updated with new types and elements to describe the different multicast substreams associated with each service. This information will include at least the multicast IP addresses at which each substream can be accessed, as well as the maximum bit rate associated with the substream. In addition, if SVC is used, there must be a mechanism to differentiate the substream carrying the base layer from the others. This way, the UE is provided with all the network parameters necessary to receive the substreams of the BC TV service with the subscribed quality definition level.

One issue related to the use of traditional coding schemas in TISPAN IMS-based IPTV is that the reception of BC TV services is always limited by the capabilities of the UE. Therefore, if the user is associated with a standard definition device, it will not be able to access high-definition channels even if they are included in the user subscription. TISPAN specifications allow for personalizing the service selection information provided to the UE, so in this case it would not be provided with information on high-definition services. Nevertheless, the

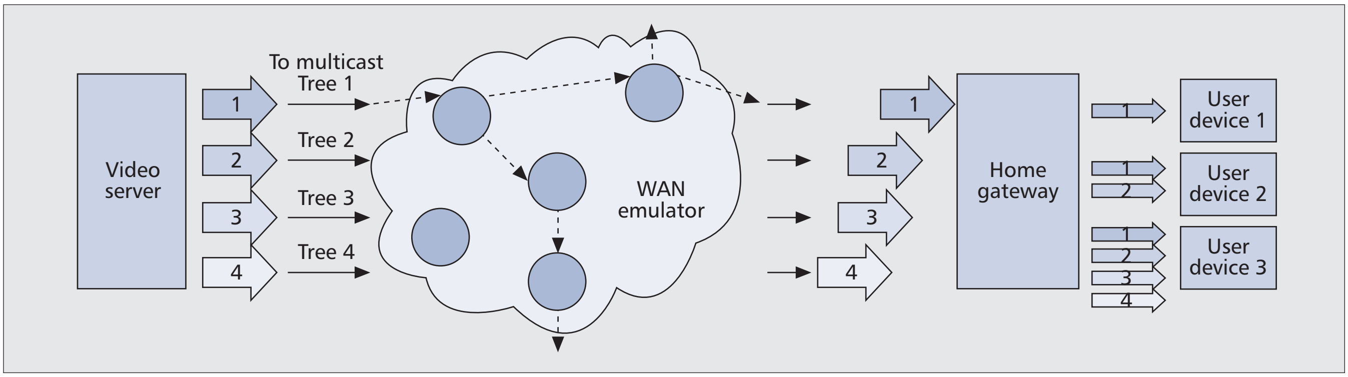

Figure 4. *Testbed schema.*

introduction of layered coding ensures that the UE can always receive every BC TV service to which the user has subscribed, no matter the quality definition level. Assuming that the UE is capable of processing at least one substream (the base layer in the case of SVC), it can always display every TV channel to the user and adjust the delivered quality according to its capabilities, just by joining more substreams (up to the subscribed quality definition level).

Regarding the session initiation procedure, it is similar to the one in Fig. 3. The difference is that in this case, the SDP offer contained in the `SIP INVITE` request carries a set of media descriptions (one for each BC TV service substream). The information included within each media description is set in a similar way as in the general case described in the previous section. The difference in this case is that the media description contains the multicast address of the substream instead of the multicast address of the BC TV service. In addition, the bandwidth requirements within the media description are defined by the largest bandwidth of all the substreams forming the BC services involved in the session. Finally, the UE joins the multicast address corresponding to each BC TV service substream.

The proposed mechanism provides more efficient usage of resources in transport networks. In TISPAN specifications, during the session initiation the bandwidth requirements are set to the largest bandwidth of all the BC services included in the session. This way, it is possible to reserve a given set of resources to cover the bandwidth demands, but only use a subset of them because the user never changes to the most bandwidth-demanding channel. By introducing layered coding, the UE initiates a session to a given BC service indicating as many media descriptions as the number of streams it wants to join. If, later on, the user changes to a higher-definition BC service, the UE can do a fast channel change simply by leaving the multicast groups of the previous BC service and joining as many multicast substreams of the new TV channel as media descriptions it previously declared. Later on, if the UE remains for some predefined time in the selected channel, the UE can modify the session by means of a `SIP UPDATE` request, introducing additional media descriptions in order to join the remaining BC TV service substreams.

Finally, in TISPAN solutions, whenever the UE receives an indication of insufficient bandwidth to initiate a BC session, it can generate a new `INVITE` request, restricting the list of BC services it intends to join and specifying a lower maximum bandwidth. This procedure may be repeated, and the session initiation may even fail if no agreement is reached. With layered coding, the UE can receive every BC service subscribed by the user (providing there is available bandwidth for one substream), and the quality perceived by the user depends on the number of substreams that may be delivered with the available bandwidth.

Multimedia Testbed

The solution described in this section has been integrated into a multimedia testbed in order to validate the proposal in a real scenario.

The main part of this testbed is the residential gateway that was initially developed within the framework of a European research and development project Multi-Service Access Everywhere (MUSE) and demonstrated in [9]. The architecture of this gateway was designed in order to support a QoS scenario by means of resource reservation and flow prioritization for different home services. The gateway can also mark packets (and understand received marked packets) when the access network can use these marks (DiffServ code point [DSCP] in the IP header, layer 2 tags, etc.) to process the packets accordingly. In the mentioned research project it was an all-Ethernet access network capable of promoting the QoS almost end to end.

The gateway was extended afterward so that it could automatically provide a call admission control and resource reservation mechanisms by intercepting and inspecting SIP messages, and was capable of being integrated in an NGN environment (focused on TISPAN architecture) also supporting IMS-enabled terminals [10].

In this trial (Fig. 4) we tested a home environment with three different devices that are able to receive three different video qualities (a smart phone, a laptop, and a high-definition television). It is important to note that there is no available implementation yet to support SVC for smart phones or televisions. In order to do the tests, we developed a generic client to generate the required SIP messages to tune a device into a TV channel as defined in the previous section.

The SIP request will indicate the specific layers of a certain channel it wants to receive that, according to the TISPAN specification, will be obtained using HTTP (the number of layers it wants depends on the desired quality and is self-limited by the quality it can support).

Whenever a device (generic client) tries to tune into a channel, the residential gateway will intercept the signaling message and analyze it, determining whether there are enough resources to receive the video or not (evaluating the available bandwidth in its different interfaces and the delay commitment that can be achieved). If the flows are allowed, resources will be reserved and the video will be prioritized when it is received. Otherwise, an indication of insufficient bandwidth is sent to the device (as if it were the core IMS) so that it can retry a more restricted subscription.

What is important here is to be able to provide the highest priority in the network (and in the gateway) to at least one layer per channel so that every device can receive it (for an SVC scenario this would be mandatory for the base

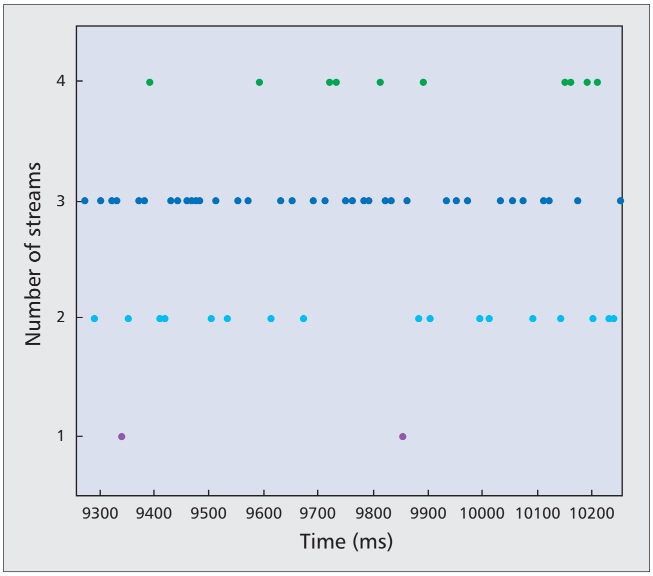

Figure 5. *Analysis of the received video quality.*

layer). In the testbed videos are split in four different layers. One of them will have the highest priority (its multicast tree will be configured in the network and in the gateway to have the highest priority); the two following ones will receive the same treatment (equal quality); and finally, there is another layer with lower priority (best effort). All the different devices will be subscribed to the highest priority layer (to guarantee some video reception); the laptop in addition will subscribe to another one and the television to all of them.

Once this process is approved (the corresponding SIP message is received), the client will launch as many video player instances as layers to be received. The video player will automatically generate the multicast signaling messages with the information provided as detailed in the previous section (the residential gateway will intercept these messages to provide IGMP snooping facilities), and the different flows will be received and analyzed in the client (if the proper codecs were available for the television, for example, the different flows would be merged into a single video).

In the testbed the access and core networks are emulated; different delay conditions are also emulated for the video streams so that their different prioritizations can be perceived. This prioritization is indicated by the video server, which marks each stream according to its priority by means of the DSCP field of the IP header.

Figure 5 shows the results of one of the analyses done in the client for the number of video streams that in a certain moment are received by the television (the client that has a subscription to every layer). Most of the time, the number of received layers is 3 and even the 4 of them, although due to the low priority assigned to the fourth layer it sometimes exhibits an unacceptable delay (of course, this all depends on buffering parameters). It is important to notice that, although using MDC all received streams in Fig. 5 will be processed to play the video, in SVC the reception of the base layer is mandatory.

Conclusion

Due to the increasing impact and deployment of IPTV nowadays, the recent definition of the IMS-based IPTV architecture may be the final push to convince network operators to massively deploy NGNs based on TISPAN recommendations. Services defined by TISPAN cover a wide range of applications demanded by end users that can easily be extended without any change to the proposed architecture, just adding the concept of layered video coding into the IPTV user profile defined by TISPAN. It has been shown that with our proposal, both end users and providers can benefit. As future work, we have plans to extend the testbed presented in this article to support the channel change, quality update, and mobility scenarios.

References

[1] ETSI TS 182 027 v. 2.0.0, " Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; IPTV Functions Supported by the IMS Subsystem," Feb. 2008.
[2] ETSI TS 123 002, "Digital Cellular Telecommunications System (Phase 2+); Universal Mobile Telecommunications System (UMTS); Network Architecture (3GPP TS 23.002 v. 7.3.1)."
[3] ETSI TS 183 063 v. 2.1.0, "Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IMS-Based IPTV Stage 3 Specification," June 2008.
[4] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-Driven Layered Multicast," *Proc. ACM SIGCOMM '96*, Aug. 1996.
[5] ITU-T Rec. H.264, "Advanced Video Coding for Generic Audiovisual Services."
[6] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Process.*, vol. 18, no. 5, Sept. 2001, pp. 74–93.
[7] B. Li and J. Liu, "Multirate Video Multicast over the Internet: An Overview," *IEEE Network*, vol. 17, no. 1, Jan./Feb. 2003, pp. 24–29.
[8] ETSI TS 102 034 v. 1.3.1, "Digital Video Broadcasting (DVB); Transport of MPEG-2 TS Based DVB Services over IP Based Networks," Oct. 2007.
[9] F. Valera *et al.*, "Demo of Triple Play Services with QoS in a Broadband Access Residential Gateway," *IEEE INFOCOM '06*, Barcelona, Spain, Apr. 2006.
[10] I. Vidal *et al.*, "Signalling Cases and QoS Management within TISPAN NGN Residential Environments," *J. Multimedia*, vol. 2, no. 2, 2007.

Biographies

IVAN VIDAL (ividal@it.uc3m.es) received his Telecommunication Engineering degree in 2001 from the University of Vigo, Spain, and his Ph.D. in telematics engineering in 2008 from University Carlos III of Madrid. His research interests include network security, multiparty services based on network multicast in the IMS, and peer-to-peer overlays for multimedia streaming in the Internet and IMS-based next-generation networks.

JAIME GARCIA received his Telecommunications Engineering degree in 2000 from the University of Vigo and his Ph.D. in telecommunications in 2003 from University Carlos III of Madrid. He is currently an associate professor at University Carlos III of Madrid having joined in 2002, and he has published over 35 papers in the field of broadband computer networks in magazines and congresses.

FRANCISCO VALERA received his Telecommunication Engineering degree in 1998 from the Technical University of Madrid and his Ph.D. in telecommunications in 2002 from University Carlos III of Madrid, where he is currently a tenured associate professor. He has published over 50 papers in magazines and conferences in the field of advanced communications. He has also has participated in scientific committees, organizations, and technical reviews for various national and international conferences.

ALEX BIKFALVI is a research assistant at the IMDEA Networks research institute and a Ph.D. student at University Carlos III of Madrid. He received a Master of Science degree in telematics engineering from University Carlos III in 2008 and a degree in electrical engineering (telecommunications) from the Technical University of Cluj-Napoca in 2006. His research interests include peer-to-peer overlays and content distribution using application-layer multicast in the Internet and next-generation networks.