

Invited Article

Standard-Compliant Content Adaptation in IPTV Systems

Truong Cong Thang¹, Jung Won Kang², Nam Pham Ngoc³, Anh T. Pham¹

¹ The University of Aizu, Japan

² Electronics and Telecommunications Research Institute, Korea

³ Hanoi University of Science and Technology, Vietnam

Correspondence: Truong Cong Thang, thang@u-aizu.ac.jp

Manuscript communication: received 31 January 2012, accepted 1 May 2012

Abstract– IPTV, which is based on the universal IP infrastructure, has the inherent nature of heterogeneity in terms of content, networks, terminals, and users. An important solution to cope with such heterogeneity is content adaptation. This paper reviews the standardization issues related to content adaptation in IPTV standards. We first describe the basic architecture of content adaptation and its integration into the ITU-T IPTV architecture. Then typical methods of content adaptation in practical IPTV systems are discussed in detail. Especially, we highlight the standard metadata tools that are potential to support adaptation methods within ITU-T IPTV architecture. Some recent developments in other standard bodies are also discussed.

Keywords– IPTV, content adaptation, MPEG-M, HTTP streaming.

1 INTRODUCTION

A major goal of IPTV standards is to enable the interoperability among content providers, service providers, network operators, and terminals used by end-users [1]. However, the development of technologies has resulted in a wide variety of content formats, protocols, devices. Moreover, different users (or user groups) may have different preferences or requirements on perceiving content. In this context, content adaptation is an important solution to enable seamless end-to-end delivery of contents [2]. For example, reducing video spatial resolution is needed to support terminal with small display size and adjusting colour or contrast of video contents may be helpful to users having visual impairments. Besides, when a content item is adapted, the metadata associated with that content item may be adapted accordingly. For example, when a provider converts in advance a video from MPEG-2 format to H.264 format, the metadata element describing the video format should be changed into H.264 as well.

In this paper, we review the standardization issues related to content adaptation in IPTV standards, with the focus on ITU-T IPTV. We first describe the basic architecture of content and metadata adaptation and its integration into the ITU-T IPTV architecture. Different cases, corresponding to different locations of adaptation within the architecture, are discussed. The paper then highlights typical adaptation methods which should be considered in practical IPTV systems. For interoperability purpose, we describe a variety of standard metadata that support such adaptation methods. A use case is explained in detail to illustrate content adaptation in

ITU-T IPTV architecture. Finally, recent developments in other standard bodies are briefly mentioned.

2 BASIC ARCHITECTURE OF CONTENT AND METADATA ADAPTATION

The basic architecture of content and metadata adaptation is shown in Figure 1. Three key modules of this architecture are decision engine, content adaptation executor, and metadata adaptation executor. The decision engine (also called Adaptation Decision Taking Engine-ADTE) takes as input the necessary metadata (or descriptions) of resource constraints, user preferences, content characteristics, etc. Based on such information, the decision engine decides adaptation instructions which are in turn the inputs of the content adaptation executor. The content adaptation executor may contain various sub-modules to carry out different adaptation operations according to the decision engine's instructions. A large amount of metadata which are necessary for the decision engine has been standardized by several standard bodies (e.g., MPEG, ETSI). The metadata associated with the input contents may be adapted by the metadata adaptation executor. Essentially, the modifications to metadata reflect the adaptation of contents. This process can be according to the resulting characteristics of adapted contents or directly based on the instructions of the decision engine.

In practice, the actual adaptation of content (as well as metadata) can be executed offline or online. In the offline case, different content alternatives are created in advance, and then an appropriate content alternative

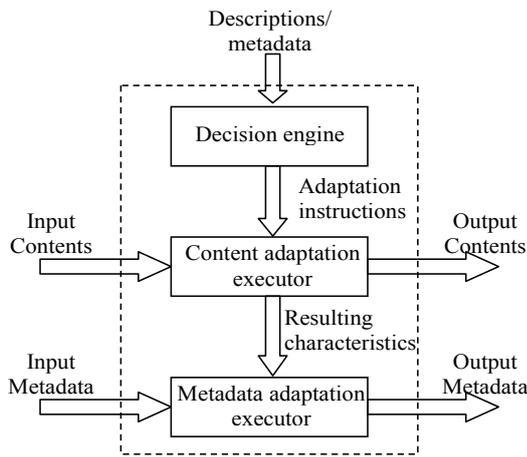


Figure 1. Basic blocks of content and metadata adaptation.

will be selected according to adaptation instructions. In the online case, the appropriate content alternative is created on the fly. Yet, in both offline and online cases, the decision engine still needs to make decisions in an effective and efficient manner.

3 ADAPTATION IN ITU-T IPTV ARCHITECTURE

The (functional) architecture of ITU-T IPTV standard is divided into functional groups (rounded rectangular boxes) and functional blocks (rectangular boxes) as seen in Figure 2. The main functional groups are Content Provider Functions, Management Functions, Application Functions, Service Control Functions, Content Delivery Functions, Network Functions, and End-User Functions. Detailed description of these functions can be found in [3]. In the following, the possible locations of content/metadata adaptation in the architecture are discussed.

3.1 Content Adaptation in ITU-T IPTV Architecture

Content Adaptation at Content Provider Functions: In practice, content providers may envisage the potential types of terminals, to which their contents will be delivered. Then, they may adapt an original content item (e.g., a classic film) into different versions, each specific to one type of terminal (e.g., HDTV, mobile phone). Using content adaptation at Content Provider Functions has an advantage that content providers can control how their contents will be displayed at end-user terminals.

Content Adaptation at Application Functions: At Application Functions, the ADTE can be considered as a part of the Content Processing Control Functional Block and Adaptation Executor as a part of the Content Pre-processing Functional Block (highlighted in pink). Content adaptation could be enabled/disabled by Content Management Functional Block using the DRM rights from content providers. When different content alternatives are already prepared by a content or service provider, Content Pre-processing Functional Block may simply take one content alternative selected

by Content Processing Control Functional Block (i.e. offline case). On the other hand, the Content Pre-processing Functional Block may adapt the original content on the fly (online case).

Content Adaptation at Content Delivery Functions: The Adaptation Executor module can be further deployed as a function of Content Processing Functional Block (highlighted in pink) which is also under the control of the Content Processing Control Functional Block. Content adaptation at this functional block can be on the fly and complementary to the adaptation at Content Pre-processing Functional Block. For example, a video content can be down-scaled at Content Pre-processing Functional Block to meet terminal display size and additionally its bitrate can be dynamically adjusted at Content Processing Functional Block to meet bandwidth variations of a wireless link.

Content Adaptation at End-User Functions: If an end-user has the right to adapt the received content, applying content adaptation at IPTV Terminal Functions would be beneficial in some cases. For example, video frame rate down-scaling by a terminal may be effective to meet the terminal's power limitation while reducing the workload of service provider.

3.2 Metadata Adaptation in IPTV Architecture

In ITU-T IPTV architecture, metadata adaptation can be carried out by the Metadata Processing Functional Block (highlighted in yellow). Similarly to the basic architecture, metadata adaptation in Metadata Processing Functional Block would reflect the resulting characteristics/information of adapted contents in Content Pre-processing Functional Block and Content Processing Functional Block. Besides, as content adaptation can be employed by a content provider, the corresponding metadata adaptation could be applied at Content Provider Functions as well.

4 ADAPTATION METHODS AND SUPPORTING METADATA

4.1 Content Adaptation Methods

This Section describes general adaptation methods which are typical for IPTV. It is obvious that these adaptation methods can be combined in different manners to produce appropriate contents for user.

Content Scaling: This method reduces some resource attributes (for example bitrate, display size, frame rate) of contents without changing their coding formats. In general, coding formats can be divided into two categories, non-scalable formats and scalable formats. Most of existing coding formats are non-scalable, which have very limited scalability. Recently, some scalable coding formats (e.g., JPEG2000, SVC) have been standardized to allow simple truncation of a coded bitstream.

Format Conversion: This is the transformation of contents from one coding format to another coding format of the same modality (e.g., MPEG-2 to H.264 of video

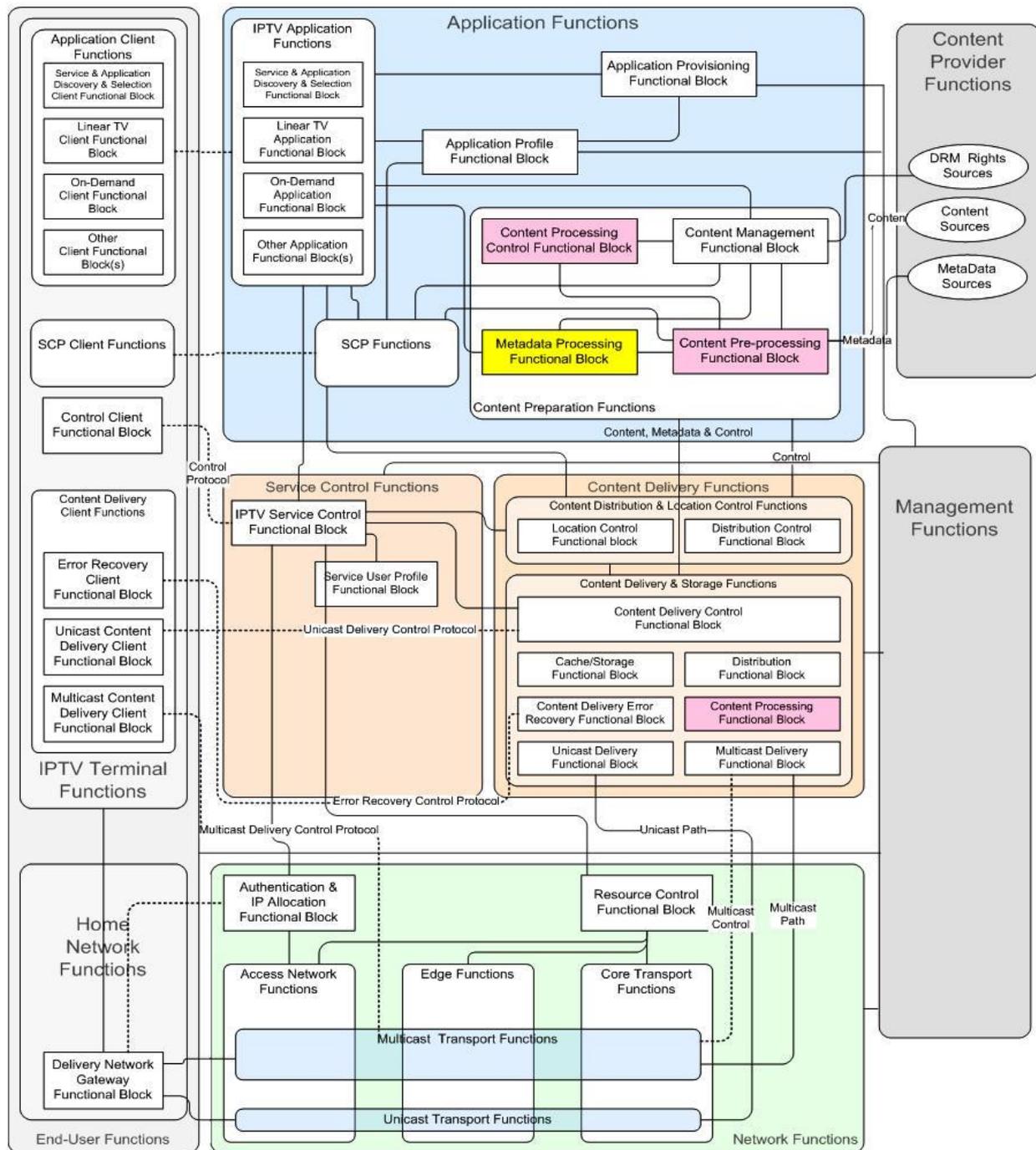


Figure 2. ITU-T IPTV Functional Architecture.

modality). This method enables seamless content delivery to end-systems having different codecs of the same modality type.

Modality Conversion: When the original modality of the content cannot be supported or decently presented at an end-system, the content should be converted to another modality which is the most suitable to the user. For example, when the adapted video quality at low bitrate is poor, the provider may send a message asking whether the user want to perceive the content as a sequence of important images.

Transcripting: For content protection, different terminals may support different cryptographic (encryption) algorithms. Similar to format conversion, this

adaptation method converts cryptographic algorithms associated with contents, facilitating the seamless content delivery between end-systems supporting different DRM standards.

Content Enhancement: With this method, contents may be enhanced (or enriched) for some users who may have disabilities or some special preferences in consuming the content. For example, the contrast or colour of video content might be strongly adapted for visually impaired users. Usually, this method is not employed to meet the resource constraints (e.g., limited connection bitrate), but to improve the quality of experience (QoE) of users.

Content Restructuring: When the content subject to adaptation is a composition of multiple objects (e.g., a web document with video, image, text objects), the content can be adapted by restructuring the content (e.g., dropping/grouping some objects). Usually, this method involves modifications of the presentation of the adapted content (e.g., using a style sheet).

4.2 Standard Metadata Supporting Content Adaptation

To support content adaptation, there needs to be information about different entities involved in a content delivery chain, including content, networks, terminals, and users. Various standard metadata have been described in the high-level specification of metadata for IPTV services [4]. In this part, we additionally highlight standard metadata tools that could potentially support the above adaptation methods. Note that, for system standards such as IPTV, ITU-T mostly refers to protocols and media/metadata formats defined by other standard bodies (e.g., MPEG, ETSI, IETF).

Given the information about adaptation constraints to be met (e.g., connection bitrate, display size), it is up to the service provider to appropriately (down-)scale the contents before sending to the terminal. This process can be facilitated by some metadata that describe the “adaptation behaviour” of contents. One powerful metadata tool for this purpose is AdaptationQoS (Clause 9.3 of ISO/IEC 21000-7 [5]). Essentially, AdaptationQoS describes the relationship between adaptation operations, the resulting resource attributes, and utilities (i.e. quality) of the adapted contents. Based on this kind of information, the decision engine will decide the best adaptation operation for a specific situation. Usually, the decided adaptation operation is the very adaptation instruction of Figure 1. Detailed discussion of AdaptationQoS for video content adaptation can be found in [6].

When applying modality/format conversion, one original modality/format could be possibly converted to several destination modalities/formats. It is necessary for users or providers to specify their preference on the order of modalities (formats) to which an original modality (format) should be converted. One metadata tool for this purpose is Conversion Preference [5]. Using Conversion Preference, a user/provider can specify the original modality/format and destination modality/format of a conversion and the “preference” for that conversion [7]. Each modality/format can be identified by a standard term defined by Content CS or some Coding Format CS of MPEG-7. Essentially, this tool could be applied to transcrypting as well.

To support content enhancement for auditory or visual accessibility, the symptoms of auditorily/visually impaired users should be described in an interoperable manner. Two possible tools are Auditory Impairment and Visual Impairment (Clauses 6.4.14 and 6.4.15 of [5]). These tools are also included in Accessibility Information metadata of ETSI TS 102 822-3-3. The Auditory Impairment describes the hearing threshold

shifts at ten frequencies in the range of 0-8 KHz. Meanwhile, Visual Impairment tool includes two kinds of visual problems: colour vision deficiency and low vision. Colour vision deficiency may describe a user’s deficiency (symptom) in perceiving red/green/blue colours. The low vision impairment of a user is described by seven basic symptoms relating to visual acuity and vision field loss.

Content restructuring in practice is a complex process, which is highly dependent on the particular contents. Due to this fact, there are few effective description tools which can support different cases of content restructuring. One possible tool is Presentation Priority Preference (Clause 6.4.13 of [5]). Basically, this tool describes the priorities of different objects of a complex (or structured) content. Based on the priority values, an adaptation system may decide to drop unimportant objects of the content.

5 ADAPTATION USE CASE

In this section, a use case is presented to illustrate the use of content adaptation in IPTV. The procedural flows in Figure 3 represent the high-level sequence of flows for a Video on Demand application that uses content adaptation at Application Functions and Content Delivery Functions. The parts related to content adaptation are highlighted in pink. Here it is assumed that provisioning and network attachment have been completed. The detailed steps in a service session are as follows.

- 1) IPTV Terminal Functions (ITF) connects to and interacts with the Application Functions (AF) to select the content item that the customer wishes to receive.
- 2) The AF connects to the Content Delivery and Device Control Functions (CDSF), to establish the authority for the ITF to consume the content.
- 3) Based on end-user profile and content metadata, the AF prepares the content (including content adaptation/selection if needed) to be delivered to the ITF.
- 4) The AF returns the URL of the CDSF and (adapted) content item.
- 5) ITF connects to the CDSF to request content delivery.
- 6) The CDSF determines which delivery function has the requested content and allocate that function.
- 7) The CDSF requests the allocation of the network resources needed to support the delivery of content to the ITF.
- 8) The ITF issues a play request.
- 9) The CDSF streams the content to the ITF.
- 10) The AF monitors the delivery quality by periodically requesting current status (e.g., bandwidth, error rate, visual quality) of ITF and CDSF.
- 11) The AF makes decision on content adaptation based on the current status.
- 12) If adaptation is needed, the AF sends adaptation instructions to CDSF.

- 13) CDSF executes the instructed adaptation operation and sends adapted content to ITF.
- 14) At the end of the session, the ITF closes the streaming session. Then the CDSF releases the delivery resources and confirms that the session is closed.

Note that Steps 10, 11, 12, 13 above could be repeated periodically during the whole session. Regardless of the nature of adaptation in Steps 3, 11, 12, 13, which could be done offline or online, the important in these steps is the metadata that is employed by and signaled between the functional entities.

6 RECENT DEVELOPMENTS

Besides ITU-T, other standard bodies also target at IPTV in somewhat different approaches. In this Section, we will highlight two new standards, namely DASH (Dynamic and Adaptive Streaming over HTTP) [8] and MPEG-M (also known as Multimedia Service Platform Technologies) [9].

6.1 Dynamic and Adaptive Streaming over HTTP

Dynamic and Adaptive Streaming over HTTP (DASH) is a new standard co-developed by MPEG and 3GPP. As implied by its name, the most important feature of this standard is the adaptivity (to network/terminal status) in which content adaptation is definitely needed. So far, IPTV services have been mostly provider-driven, where system intelligence locates at the provider's side. On the contrary, DASH is based on the client-driven model and takes advantage of the abundant availability of existing Web platform, including Web servers and Content Delivery Networks (CDN).

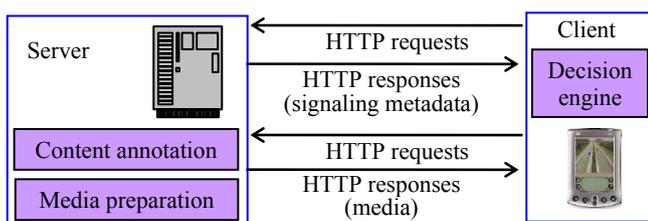


Figure 4. HTTP streaming architecture.

The basic architecture of adaptive HTTP streaming is shown in Figure 4 [10]. Here, the content annotation module provides knowledge (or metadata) about the content, ranging from semantic level (e.g., genres) to physical level (e.g., bitstream structure/characteristics). The media preparation module provides tools for media adaptation/transcoding, packetization, encapsulation, etc., so that the media could be efficiently delivered to the client. The metadata provided by content annotation module will be requested by the client. Based on the metadata and status of terminal/networks, the decision engine at the client makes decision on which/when media parts are downloaded. In essence,

MPEG DASH specifies the formats of signaling metadata and media exchanged between clients and servers. With conventional streaming, signaling metadata and media are usually delivered by different protocols (e.g., by RTSP and RTP respectively) [11]. Meanwhile, in HTTP streaming, both signaling metadata and media are usually delivered by HTTP protocol. The metadata sent from a server to a client is called Media Presentation Description (MPD) in DASH. Content may be composed of one or more content component (e.g., video, audio). The content could be divided into one or more temporal chapters (or periods). Further, each alternative (also called representation) of a content component could be divided into media segments. The client will process the metadata for a period, and request appropriate segments of that period. It then gets metadata for next period, and the whole process is started over. For the purpose of adaptivity, the signaling information could be generally divided into the following categories [12].

General information: this information is the high-level description of content and could be divided into 2 levels. The first is content-level information that provides the common description of a whole presentation, such as available time, duration, minimum initial buffering. The second is period-level information that provides the common description of a period of a presentation, such as start time, duration, unique identifier, etc.

Adaptation information: this information describes the characteristics of each alternative/representation of a content item, such as bitrate, resolution, quality. An alternative could be physical, i.e. created in advance, or virtual, i.e. to be created on the fly. Based on the information of alternatives, the client will select a segment of appropriate alternative, and thus supporting the adaptivity to contexts of terminal and networks.

Mapping information: this information describes the locations to retrieve actual media data. Depending on specific cases, different alternatives could have the same or different locations.

Client request: this is a special kind of signaling information sent from a client to a server, which shall conform to the format of HTTP 1.1. As shown in the below figure, the parameters to be requested by the client will be derived from the information of the above categories.

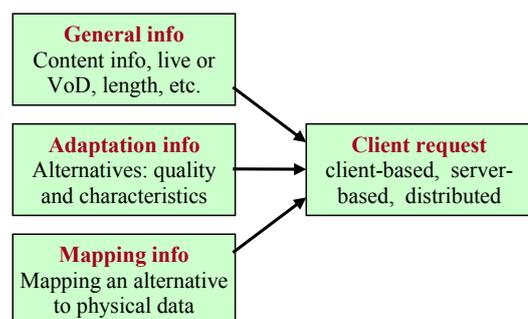


Figure 5. Categories of signaling metadata.

For efficiency in managing/processing the metadata,

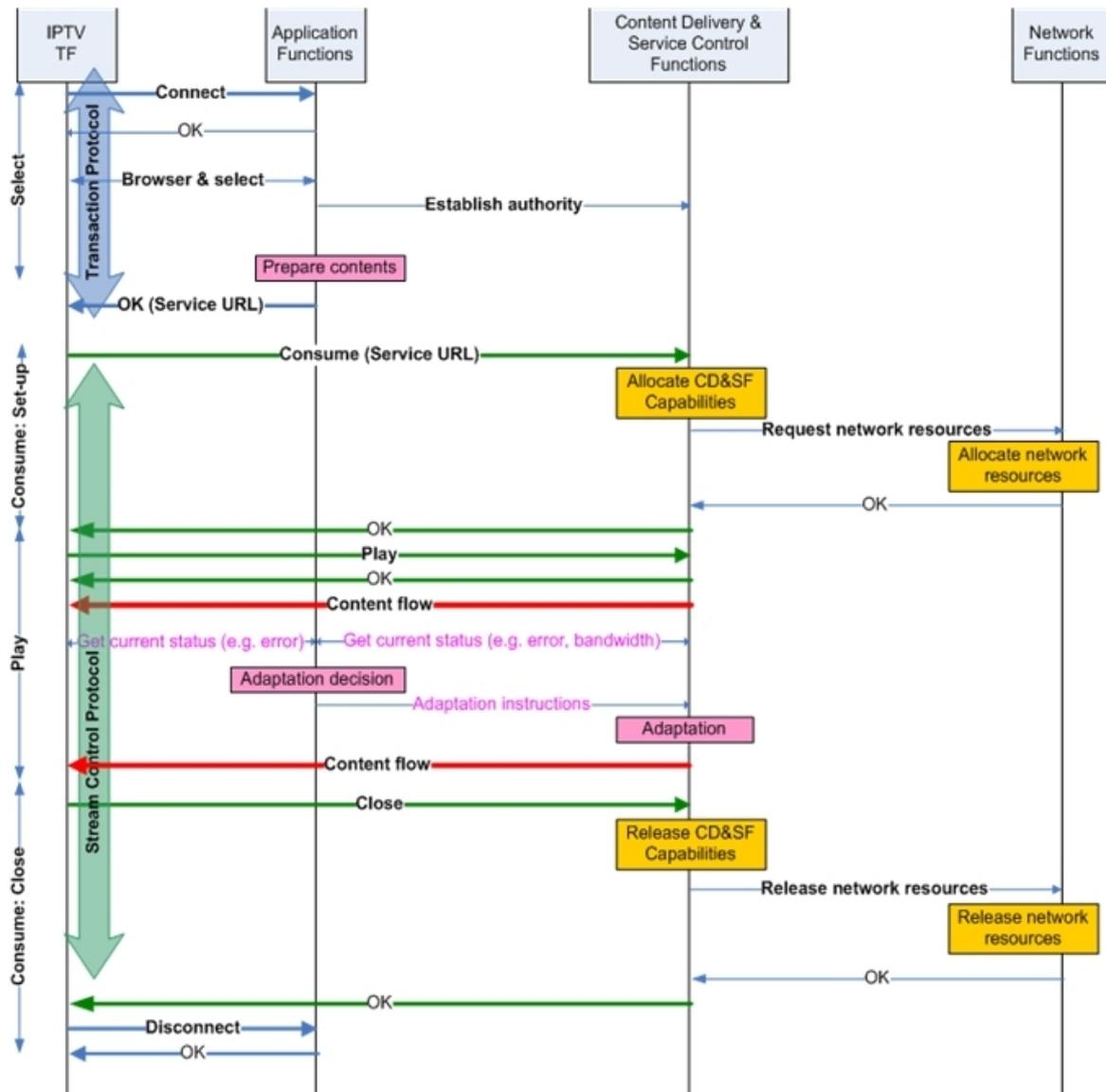


Figure 3. A session of IPTV VoD service with content adaptation.

a metadata document may be fragmented into different parts. The separation of metadata parts also enables efficiency in storage and delivery. For example, during a session, content-level metadata could be sent once, and only period information is periodically loaded. More information about the structure and basic concepts of DASH could be found in [8], [13].

6.2 MPEG-M

Currently, MPEG group is going to finalize MPEG-M standard, which is originally targeted at the objectives of Advanced IPTV Terminal (AIT) project [14]. Different from the rather fixed architecture of ITU-T IPTV, MPEG-M provides a flexible and extensible middleware based on a large number of “elementary services” with standard protocols and Application Programming Interfaces (APIs). In fact, MPEG-M, also referred to as Multimedia Service Platform Technologies, is a set of 6 specifications (or parts) [9], namely:

- Part 1 - Architecture,
- Part 2 - Application Programming Interface,

- Part 3 - Conformance and Reference Software,
- Part 4 - Elementary Services,
- Part 5 - Service Aggregation, and
- Part 6 - Usage Guidelines.

. These specifications not only specify high-level protocols and message formats, but also leverage many existing technologies of MPEG (e.g., those in MPEG-4, MPEG-7, MPEG-21) in a common framework.

The most important among these six parts of MPEG-M is Part 4, which defines a table of elementary services or the building blocks of any applications. Here, each elementary service is an operation (Deliver, Search, Process, etc.) applied to a specific type of entity (Content, Contract, Device, etc.). In MPEG-M, content adaptation belongs to a type of elementary service called Process Content. Also, in the elementary service of Request Content, the content provider may select or adapt a content alternative to meet the client’s context (e.g., display size, connection throughput). MPEG-M enables the fast and easy deployment of various business models (like IPTV) in a standard manner,

where each player/provider can take part in the value chain by combining/optimizing a number of elementary services. In such business models, content adaptation could be provided by some special providers, for example a portal provider that adapts and delivers multimedia content on Internet to any mobile devices.

7 CONCLUSIONS

In this paper, we have reviewed standardization issues related to content adaptation in IPTV. We described the basic architecture of content adaptation and its integration into the ITU-T IPTV functional architecture. Typical methods of content adaptation in IPTV systems, including content scaling, format/modality conversion, transcoding, content enhancement, and content restructuring, were discussed in detail. For interoperability purpose, we highlighted a variety of standard metadata that support such adaptation methods. Also, the adaptivity feature in two recent developments, namely DASH and MPEG-M, was briefly discussed. Among these two new standards, DASH is targeted at the urgent demand of Internet video streaming while MPEG-M is expected to provide a common and comprehensive platform for different entertainment and communication services in the future.

REFERENCES

- [1] S. Wright, S. Jones, and C. S. Lee, "IPTV systems, standards, and architectures: Part I," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 69–96, 2008.
- [2] T. C. Thang, *Content and Metadata Adaptation for IPTV Services*, ITU-T Std. Working Document SG16 Doc. TD-62, 2009.
- [3] *IPTV Architecture*, ITU-T Std. Working document FG-IPTV DOC-0084, 2007.
- [4] *High-level specification of metadata for IPTV services*, ITU-T Std. Recommendation H.750, 2008.
- [5] *Information technology - Multimedia framework (MPEG-21) - Part 7: Digital Item Adaptation*, ISO/IEC Std. 21000-7, 2005.
- [6] T. C. Thang, J. G. Kim, J. W. Kang, and J. J. Yoo, "SVC adaptation: Standard tools and supporting methods," *Signal Processing: Image Communication*, vol. 24, no. 3, pp. 214–228, Mar. 2009.
- [7] T. C. Thang and Y. M. Ro, "Modality conversion: Toward the semantic frontier of UMA," in *Multimedia Transcoding in Mobile and Wireless Networks*, A. M. A. Ahmad and I. K. Ibrahim, Eds. Information Science Reference, 2008.
- [8] *Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segment formats*, ISO/IEC Std. IS 23009-1, 2012.
- [9] P. Kudumakis, X. Wang, S. Matone, and M. Sandler, "MPEG-M: Multimedia service platform technologies," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 159–163, 2011.
- [10] T. C. Thang, J. W. Kang, and A. T. Pham, "Quality-guided adaptivity for audiovisual content streaming by MPEG DASH," in *Proc. IEEE Int. Conf. Consumer Electronics (ICCE 2012)*, Jan. 2012, pp. 94–95.
- [11] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the internet: approaches and directions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, 2001.
- [12] T. C. Thang, J. Y. Lee, J. W. Kang, S. J. Bae, S. Jung, and S. T. Park, *Signaling Metadata for Adaptive HTTP Streaming*, ISO/IEC Std. JTC1/SC29/WG11 m17771, July 2010.
- [13] T. Stockhammer, "Dynamic adaptive streaming over HTTP - standards and design principles," in *2nd Annual ACM Conf. Multimedia systems (MMSys 2011)*, California 2011.
- [14] *Context and Objectives for Advanced IPTV Terminal (AIT)*, ISO/IEC Std. JTC1/SC29/WG11 N11230, 2009.



Truong Cong Thang received the B.E. degree from Hanoi University of Technology, Vietnam, in 1997 and Ph.D degree from KAIST - IT Convergence Campus (KAIST-ICC), Korea, in 2006. From 1997 to 2000, he worked as an engineer in Vietnam Post and Telecom (VNPT). From 2007 to 2011, he was a Member of Research Staff at Electronics and Telecommunications Research Institute (ETRI), Korea. He was also an active member of Korean delegation to standard meetings of ISO/IEC and ITU-T from 2002 to 2011, specializing in IPTV, MPEG-4, MPEG-21, DASH, and MPEG-M. Since 2011, he has been an Assistant Professor of University of Aizu, Japan. His research interests include multimedia networking, image/video processing, content adaptation, IPTV, and ISO/IEC/ITU standards.



Jung Won Kang received the BS and MS degree in Electrical Engineering in 1993 and 1995, respectively, from the Hankuk Aviation University, Seoul, South Korea. She received the Ph.D degree in Electrical and Computer Engineering in 2003 from Georgia Institute of Technology, Atlanta, GA. Since 2003, she has been a Senior Member of Research Staff in Broadcasting and Telecommunications Convergence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Korea. Her research interests are in the areas of video signal processing, video coding, and video adaptation.



Pham Ngoc Nam received B. Eng. degree in Electronics and Telecommunications from Hanoi University of Science and Technology (Vietnam) and M.Sc. degree in Artificial Intelligence from K.U. Leuven (Belgium) in 1997 and 1999, respectively. He was awarded a Ph.D. degree in Electrical Engineering from K.U. Leuven in 2004. Since 2004, he has been working at Hanoi University of Science and Technology, Vietnam. His research interests include QoS management at end-systems for multimedia applications, reconfigurable embedded systems and low-power embedded system design.



Anh T. Pham received the B.E. and M.E. degrees, both in Electronics Engineering from the Hanoi University of Technology, Vietnam in 1997 and 2000, respectively, and the Ph.D. degree in Information and Mathematical Sciences from Saitama University, Japan in 2005. From 1998 to 2002, he was with the NTT Corp. in Vietnam. Since April 2005, he has been on the faculty at the University of Aizu, where he is currently an associate professor at the Computer Communications Laboratory, the School of Computer Science and Engineering. His present research interests are in the areas of computer networking, optical communications, and spread spectrum technique. Dr. Pham received Japanese government scholarship (MonbuKagaku-sho) for Ph.D. study. He also received Vietnamese government scholarship for undergraduate study. Dr. Pham is senior member of IEEE. He is also member of IEICE and OSA.