Challenges toward Adaptive Behavior of Distributed Multimedia Systems

Hermann Hellwagner
Department of Information Technology (ITEC) · Klagenfurt University, Austria
hermann.hellwagner@itec.uni-klu.ac.at

Abstract

The multimedia community is pursuing, among others, the vision of Universal Multimedia Access (UMA). UMA denotes the concept that any multimedia content should be available anywhere, anytime, on any device, tailored to the user's needs and preferences, accessible for the user in a transparent and convenient way. Key to achieving this vision is to realize collaborative adaptive behavior of the involved distributed multimedia system components (server, media-aware network elements like proxies or gateways, and clients), based on intense metadata exchange and multimedia content negotiation, adaptation, or personalization.

This paper outlines the key challenges and the state of the art in achieving such adaptive behavior. The major challenges have been tackled recently and many of the building blocks of UMA have become or are becoming available from standardization groups, which are instrumental in this area to ensure interoperable use of the media and metadata items. However, making use of these standards in practical multimedia systems raises additional issues, both of conceptual nature and in terms of implementation. Based on years of contributions to ISO/IEC MPEG standardization efforts and of research work into adaptive multimedia systems, we will give an overview of these challenges, discuss the state of the art, and introduce an emerging principled solution for format-independent multimedia content adaptation.

1. Introduction

Universal Multimedia Access (UMA) has been a research topic of the multimedia community for many years. UMA denotes the vision that any multimedia content should be available anywhere, anytime, on any device, tailored to the user's needs and preferences, accessible for the user in a transparent and convenient way [1][2][3]. The primary goal of UMA is to offer the user(s) the best quality of service (QoS) or experience (QoE) in a given multimedia consumption situation.

Achieving UMA is a major and even growing challenge. Difficulties stem from basically four sources [3], as illustrated in Figure 1:

- The multitude of multimedia content formats and of rich content offerings.
- A potentially wide spectrum of user needs, preferences, and possible impairments.
- The diversity of end devices used for content presentation.
- The heterogeneity and dynamicity of the networks over which the multimedia content is transmitted.

There are additional concerns, e.g., Digital Rights Management (DRM), but the focus of our discussions will be on those mentioned above.

In addressing these problems, a recent focus of the multimedia research and standardization communities has been on concepts and mechanisms enabling *collaborative adaptive behavior* of the involved distributed multimedia system components (servers, media-aware network elements like proxies or gateways, and clients), as also indicated in Figure 1. The corresponding "building blocks" of a UMA-enabled system can therefore be summarized as follows [3]:

- Scalable media encoding and decoding formats as well as means for describing rich, multimodal multimedia "scenes".
- Metadata describing important properties of the multimedia content, e.g., the available content variants or adaptation options.
- Metadata describing the delivery and usage context, i.e., the network situation, device properties, the user's preferences and constraints, and natural environment characteristics.
- Exchange of this metadata as well as content selection, adaptation, conversion and/or personalization mechanisms, with the latter also pertaining to graphical user interfaces and applications.

Interoperability of the metadata and solutions realizing these "building blocks" is of crucial concern. For instance, a network element like an adaptation-capable proxy server must be able to properly deal with the metadata describing the multimedia content, user preferences, network conditions and device capabilities,

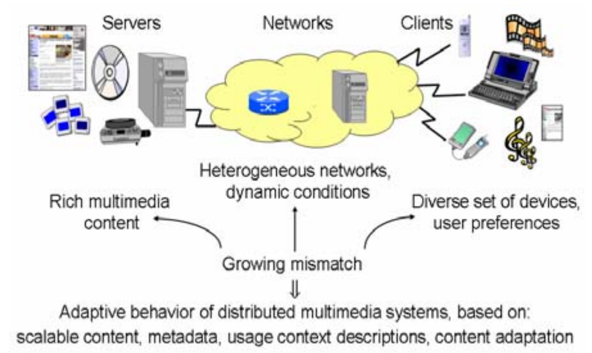


Figure 1. Challenges and concept of UMA

and must produce a content variant that is conforming to a multimedia format or language that the user's device can cope with and that is tailored to the current usage context and the user's needs. Standards and standardization bodies therefore play a central role in bringing UMA closer to reality.

Based on years of contributions to ISO/IEC MPEG (Moving Picture Experts Group) standardization efforts and a number of research projects into realizing adaptive multimedia systems, we discuss in this paper the major challenges and review current solutions toward adaptive behavior of the involved multimedia system components. We briefly review relevant multimedia (metadata) standards in Section 2. In Section 3, we exemplarily address several questions that emerge when realizing such behavior toward UMA in a fully functional multimedia adaptation solution. We also describe and discuss an emerging principled solution for format-independent multimedia content adaptation. Section 4 provides concluding remarks.

2. Multimedia (Metadata) Standards Supporting UMA

In order to address the challenges involved in the UMA vision while ensuring interoperability, ISO/IEC and other standardization groups have been working on developing content formats, metadata specifications, and frameworks suited as "building blocks" of UMA systems. Since we have contributed to the corresponding efforts of *ISO/IEC MPEG* for several years, we summarize below the major results of ISO/IEC in this area (Table 1).

The most relevant technologies or specifications for UMA are the normative descriptions defined in the MPEG-7 [4][5] and MPEG-21 [6][7] families of standards as well as recent scalable content formats, most notably the Scalable Video Coding (SVC) extensions to the H.264 / MPEG-4 AVC coding standard [8][9]. Due to space restrictions, only a very brief introduction can be given here. A more detailed discussion of these relevant metadata standards MPEG-7 and MPEG-21 is given in [3] and in the specific literature.

While the overall goal of the *MPEG-7* standard, the *Multimedia Content Description Interface*, is content description (i.e., metadata) to enable fast and efficient indexing, searching, retrieval and filtering of multimedia material, the scalable or adaptive delivery of multimedia (in other words, UMA) is addressed by several Description Schemes (DSs) as well. For instance, content variations can be described, i.e., variants of a source content such as a low-resolution image version, a summary of a video, or a text transcript of audio material. These descriptions can include a fidelity value specifying the quality change of the variation w.r.t. the source, and the type of the variation such as spatial reduction, summary, or modality conversion. Other DSs like media information descriptors, user preferences, transcoding hints, semantics descriptions, space and frequency views, and summaries provide information for adaptation tools to perform context- and utility-aware modifications of multimedia contents both on a low, syntactical level (changing the original content signal) and on a higher, semantic level (modifying or extracting the content in a way such as to create particular utility for the content consumer).

MPEG-21, the *Multimedia Framework*, aims to provide a comprehensive open framework providing technologies for the entire multimedia content delivery chain including content creation, production, transmission, personalization, consumption, presentation, and trade. A key concept in MPEG-21 is the Digital Item (DI), a structured digital object with a standard representation (a flexible XML "container"), identification, and associated metadata. Adaptation of multimedia resources for transparent use of content on various devices and networks is specifically covered by Part 7 of the MPEG-21 standards family, entitled *Digital Item Adaptation (DIA)* [6][10][11]. DIA is instrumental for UMA because it broadly defines normative descriptions of the multimedia usage context (Usage Environment Descriptions, UEDs), i.e., user preferences and impairments, device properties, network characteristics, and natural environment (see also Table 1). This supports *device independence*, i.e., delivery and consumption of multimedia content on various end user devices, enabled by content adaptation parameterized by the UEDs. Another central element of DIA, *coding format independence*, refers to the concept that adaptation of (scalable) multimedia content can be achieved in a coding format agnostic manner. This concept will be addressed in Section 3.

UMA Building Blocks	**Supporting ISO/IEC Technologies/Specifications**
Rich, multimodal, scalable multimedia content	JPEG 2000 MPEG-4 scalable A/V codecs MPEG-4 scene descriptions: BIFS, LASeR Scalable extensions (SVC) to H.264/MPEG-4 AVC
Content-related metadata	MPEG-7
Structuring and association of content and metadata	MPEG-21 Digital Item Declaration (DID)
Context-related metadata: – User preferences – Device capabilities – Network characteristics – Usage environment	MPEG-21 Digital Item Adaptation (DIA) – Usage Environment Descriptions (UEDs)
Content selection	MPEG-21 DID: `choice/selection` elements
Content adaptation: – Transcoding/scaling on the signal level – Transcoding/scaling on the semantic level – Modality conversion	Specific MPEG-7 Description Schemes (DSs), e.g., Transcoding Hints MPEG-21 DIA – Bitstream Syntax Descriptions (BSDs) MPEG-21 DIA – Amendment 1

Table 1. UMA components provided by ISO/IEC

Finally, remarkable advances have been made in creating scalable coding formats, most importantly in the field of *Scalable Video Coding (SVC)*. The H.264/SVC standard is an extension of the well-established H.264 / MPEG-4 AVC video coding format. Video data in SVC are structured as a base and possibly multiple enhancement layers along the temporal, spatial, and quality scalability axes. Scalability here means that a single “global” bitstream may contain multiple degraded versions of the media content which may be extracted easily and represent lower-quality variants of the best quality stream, but consume proportionally less storage space and network bandwidth. SVC holds the promise that video adaptation can be performed computationally inexpensively on a media server or even on network devices, probably performed in a coding format independent way as indicated above.

3. Challenges in Realizing UMA Systems

Despite this valuable basic work done by standardization groups, making use of these standards in practical multimedia systems supporting UMA raises additional issues, some of which are addressed in the following.

Does content adaptation work in real time? As demonstrated by the Personal Live Video Server (PLVS) [12][13] of our group and by many other authors, adaptation of audio-visual (A/V) content can well work in real time, even based on a comparatively expensive transcoding approach. For instance, a handful of MPEG-2 A/V streams can be transcoded to other formats in real time, using the PLVS software on an off-the-shelf PC. Using a scalable content format like H.264/SVC, recent experiments in our group have shown that a handful of SVC streams can even be adapted (scaled) in real time on an inexpensive WLAN router with very limited computational power.

Where to perform content adaptation? This question pertains to (1) on which node(s) in the multimedia delivery chain and (2) on which layer(s) of the protocol stack content adaptation should take place.

Concerning (1), the obvious and simple option is to perform adaptation on the media server, where all the media data, metadata, and adaptation options are available. The drawbacks of this approach are that the information on the dynamically changing delivery context is probably (out-)dated due to potentially long feedback delays (from the delivery context to the server) and that many unicast connections are required to specifically serve the clients. Therefore, we favor and have worked on *distributed adaptation* where the task of changing the content can also be delegated to an adaptation proxy or gateway (or even a WLAN router, as indicated above). Such a network device can react more quickly to changes in the delivery context, which may be fast and severe particularly in wireless networks. Furthermore, the adaptation component in the network can be served by a single, best-quality stream; from there, multiple singlecast or multicast streams to the clients can be established, saving resources in the network (Figure 2). However, this approach is more difficult since the metadata have to be delivered to the adaptation node in synchrony with the media data and adaptation has to be done in a piecewise fashion. A full framework and an implementation of distributed adaptation have been worked out in the European Project DANAE [14] and refined meanwhile [15].

With respect to question (2), our research on in-network adaptation nodes indicates that the implementation of an adaptation component should be done as a proxy, i.e., on application layer. Among other concerns, this is necessitated by the fact that the usual RTP communication must be kept intact after adaptation, e.g., sequence numbers of packets must be successive even if portions of a stream are discarded. A simple, stateless packet filter on network layer is not sufficient to achieve this. This even holds for the SVC stream adapting WLAN router mentioned above. However, a cross-layer approach in an adaptation proxy or gateway (or router) is beneficial in order to timely transfer information about network conditions from the lower protocol layers (most importantly, MAC layer) up to the application layer (proxy process). But since cross-layer approaches are considered not without problems since they may run counter to interoperability and lead to "spaghetti design" [16], one of our approaches is to integrate MPEG-21 descriptors into a cross-layer approach by mapping network information into MPEG-21 DIA UEDs, thus improving interoperability [17].

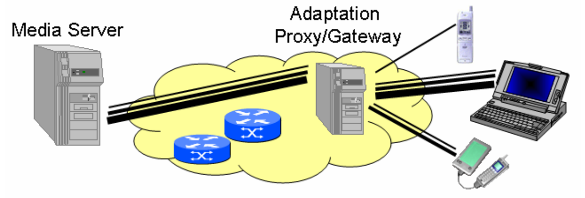


Figure 2. Possible locations for content adaptation

What and how to adapt? What to optimize? These questions mainly pertain to taking an appropriate (ideally, optimum) adaptation decision, i.e., to figure out whether to modify the video or the audio stream or both, and which of their properties to change. The objective behind this decision is to yield the best possible content quality and utility for the user under the given circumstances.

We have worked on several ways to model and perform adaptation decision taking. An evident solution is to formulate adaptation decision taking as a general *optimization* problem [18], where the objective function is to maximize content quality, and constraints are given by device, network, user, and content provider needs as well as content adaptation options and resulting qualities. All these constraints and facts are expressible using MPEG-21 DIA metadata.

A second approach is to understand and model adaptation decision taking as an *AI planning* problem [19]. The MPEG-7 content description serves as a start state and the MPEG-21 delivery context constraints to be met represents the goal state. Actions that effect state transfers are simple multimedia content adaptation operations, e.g., image re-sizing. The objective is to find, as the desired result, an adaptation plan, i.e., a sequence of adaptation steps, that takes the start state to the goal state and, thus, describes how to transform the given content step-by-step such that it eventually suits the given consumption situation.

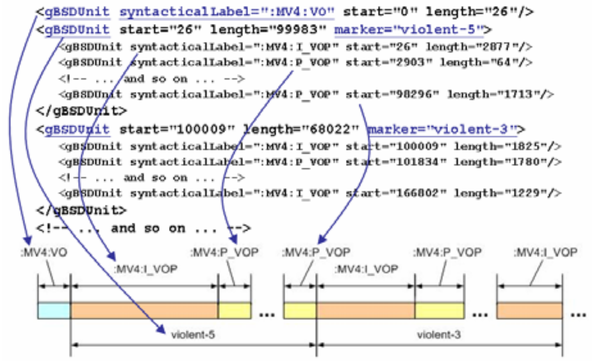
In a final approach, a model for the utility that the content delivers to the user was developed, and finding an adaptation decision is then basically to optimize that *utility* [13][20]. The utility model considers both perceptual and semantic factors. A recommendation system integrated into PLVS [13] proposes a certain content adaptation to be performed, based on basic built-in knowledge, demographic filtering and collaborative filtering, which says that persons in the same or a similar content consumption situation have preferred the proposed content variation. A feedback system enables the user to rate the content variation(s) presented. Results from subjective user tests indicate that this is a promising way to attain good user-centred adaptation.

Are there more principled solutions to multimedia content adaptation? At first glance, multimedia content adaptation necessarily appears to be a content (format) specific process. However, given the diversity of existing (and most probably, of future) coding formats as pointed out in Section 1, this seems to be prohibitive for wide deployment of content adaptation and for UMA; after all, an adaptation node would have to support every (scalable) coding format. A more principled or generic solution to content adaptation is needed.

The MPEG-21 standardization group has considered this requirement and defined description "tools" within DIA that enable coding format independent content adaptation to be achieved. So-called *Bitstream Syntax Descriptions (BSDs)* can describe (based on XML) the high-level syntax of a media bitstream, i.e., how it is organized in terms of layers, frames, slices, headers, or packets. One variant of BSDs, the *generic BSD (gBSD)* [6][11][18], contains predefined description elements that can be used to describe any multimedia content in a format independent, abstract way. In addition, gBSD provides semantically meaningful marking of bitstream segments and a hierarchical structure. These concepts enable simple types of semantic adaptations to be performed, e.g., according to violence levels (Figure 3).

Using this description-driven approach, the actual bitstream adaptation happens in two steps. In the first step, the description as exemplified above is modified (transformed), steered by the adaptation requirements of the usage and user environment (UEDs). For instance, from the gBSD of Figure 3, scenes (gBSDUnits) of a certain violence level can be discarded. This transformation can utilize standard XML techniques like XSLT or STX. In the second step, the adapted media bitstream is automatically generated on the basis of the original bitstream and the transformed BSD. Both steps can be combined for increased efficiency.

This BSD-based adaptation approach is complemented by a number of additional description "tools" that basically represent input to the adaptation decision. Besides the UEDs, information about the quality of the multimedia content (and its variations) and about additional constraints imposed by the content provider or content user can be given, e.g., that the size of an image should not be reduced below a quarter of the original size. For details and examples, we refer to [6][18].



```
<gBSDUnit syntacticalLabel=":MV4:VO" start="0" length="26"/>
<gBSDUnit start="26" length="99983" marker="violent-5">
   <gBSDUnit syntacticalLabel=":MV4:I_VOP" start="26" length="2877"/>
   <gBSDUnit syntacticalLabel=":MV4:P_VOP" start="2903" length="64"/>
   <!-- ... and so on ... -->
   <gBSDUnit syntacticalLabel=":MV4:P_VOP" start="98296" length="1713"/>
</gBSDUnit>
<gBSDUnit start="100009" length="68022" marker="violent-3">
   <gBSDUnit syntacticalLabel=":MV4:I_VOP" start="100009" length="1825"/>
   <gBSDUnit syntacticalLabel=":MV4:P_VOP" start="101834" length="1780"/>
   <!-- ... and so on ... -->
   <gBSDUnit syntacticalLabel=":MV4:I_VOP" start="166802" length="1229"/>
</gBSDUnit>
<!-- ... and so on ... -->
```

Figure 3. gBSD example

The major advantage of BSD-driven adaptation, namely the option of coding format independent content adaptation, comes at the cost of difficult generation of the BSDs (in both variants provided [6]), metadata overhead, and performance penalties. All these problems are being addressed in several projects, e.g., by compressing the BSD metadata or by applying efficient transformation methods to the BSDs, with the result that real time performance is easily achieved today.

4. Conclusion

In this paper, we presented the major challenges of adaptive behavior in multimedia systems toward the UMA vision, reviewed the pertinent results mainly of ISO/IEC standardization, and discussed remaining research problems when implementing UMA mechanisms in practical distributed multimedia systems. While best-practice solutions have been prototyped mainly within European research projects, real-world deployment of the UMA “building blocks” on a large scale is still an open issue. We assert that this situation is mainly due to UMA’s complexity, due to unclear benefits for vendors (but clear cases for proprietary solutions), and due to the lack of strategy of large-scale deployment. Yet, by developing solutions for the problems described and showing their benefits, we are contributing to bringing the UMA vision closer to reality.

5. References

[1] R. Mohan, J. R. Smith, and C.-S. Li, “Adapting Multimedia Internet Content for Universal Access”, *IEEE Trans. on Multimedia*, vol. 1, no. 1, Jan.-March 1999.

[2] A. Vetro, C. Christopoulos, and T. Ebrahimi (eds.), *IEEE Signal Processing Magazine, Special Issue on Universal Multimedia Access*, vol. 20, no. 2, March 2003.

[3] C. Timmerer and H. Hellwagner, “MPEG Standards Enabling Universal Multimedia Access”, *Proc. First Int’l. Conf. on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS 2005)*, Florence, Italy, Nov./Dec. 2005.

[4] B.S. Manjunath, P. Salembier, and T. Sikora (eds.), *Introduction to MPEG-7: Multimedia Content Description Language*, John Wiley & Sons Ltd., NJ, 2002.

[5] S.-F. Chang, A. Puri, T. Sikora, and H. Zhang (eds.), *IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on MPEG-7*, vol. 11, no. 6, 2001.

[6] I. Burnett, R. Koenen, F. Pereira, and R. Van de Walle (eds.), *The MPEG-21 Book*, Wiley, 2006.

[7] F. Pereira, J. R. Smith, and A. Vetro (eds.), *IEEE Trans. on Multimedia, Special Section on MPEG-21*, vol. 7, no. 3, June 2005.

[8] T. Wiegand, G. J. Sullivan, H. Schwarz, and M. Wien (eds.), *ISO/IEC 14496-10:2008/Amd3: Scalable Video Coding.* Int’l. Org. for Stand., 2008 (to appear).

[9] T. Wiegand, G. J. Sullivan, J.-R. Ohm, and A. K. Luthra (eds.), *IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Scalable Video Coding—Standardization and Beyond*, vol. 17, no. 9, Sept. 2007.

[10] A. Vetro and C. Timmerer, “Digital Item Adaptation: Overview of Standardization and Research Activities”, *IEEE Trans. on Multimedia*, vol. 7, no. 3, June 2005.

[11] C. Timmerer and H. Hellwagner, “Interoperable Adaptive Multimedia Communication”, *IEEE Multimedia Magazine*, vol. 12, no. 1, Jan.-March 2005.

[12] M. Prangl and H. Hellwagner, “A Framework for Personalized Utility-aware IP-based Multimedia Consumption”, *IEEE Int’l. Symp. on a World of Wireless, Mobile and Multimedia Networks (WOWMOM 2007)*, Helsinki, Finland, June 2007.

[13] M. Prangl, *Quality and Utility Modelling of Multimedia Contents for Improved Multimedia Experience*, PhD Thesis, Klagenfurt University, 2007.

[14] A. Hutter, P. Amon, G. Panis, E. Delfosse, M. Ransburg, and H. Hellwagner, “Automatic Adaptation of Streaming Multimedia Content in a Dynamic and Distributed Environment”, *Proc. Int’l. Conf. on Image Processing (ICIP 2005)*, Genova, Italy, Sept. 2005.

[15] M. Ransburg, C. Timmerer, S. Devillers, and H. Hellwagner, “Design and Evaluation of a Metadata-Driven Adaptation Node”, *Proc. 8th Int’l. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2007)*, Santorini, Greece, June 2007.

[16] V. Kawadia, P.R. Kumar, “A cautionary perspective on cross-layer design”, *IEEE Wireless Communications Magazine*, vol. 12, no. 1, Feb. 2005.

[17] I. Kofler, C. Timmerer, H. Hellwagner, and T. Ahmed, “Towards MPEG-21-based Cross-layer Multimedia Content Adaptation”, *Proc. 2nd Int’l. Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, London, UK, Dec. 2007.

[18] C. Timmerer, *Generic Adaptation of Scalable Multimedia Resources*, PhD Thesis, Klagenfurt Univ., 2006.

[19] K. Leopold, *Knowledge-based Multimedia Adaptation*, PhD Thesis, Klagenfurt University, 2006.

[20] M. Prangl, T. Szkaliczki, and H. Hellwagner, “A Framework for Utility-based Multimedia Adaptation”, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 6, June 2007.