INTEROPERABLE ADAPTIVE MULTIMEDIA COMMUNICATION

Christian Timmerer and Hermann Hellwagner

Department of Information Technology (ITEC), Klagenfurt University, Austria
{christian.timmerer, hermann.hellwagner}@itec.uni-klu.ac.at

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT

Department of Information Technology (ITEC)
Klagenfurt University
Technical Report No. TR/ITEC/05/1.01
January 2005

Interoperable Adaptive Multimedia Communication

Christian Timmerer and Hermann Hellwagner
Klagenfurt University

The desire to gain access to rich, multimedia-based information anywhere, anytime grows enormously. Web sites without multimedia-enriched information—including text, images, audio, and video samples—might no longer be appropriate to satisfy today's users. The appetite for accessing multimedia-based content via almost any type of terminals—seamlessly over dynamic and heterogeneous networks, independently of location or time—is similarly growing. To achieve such access, the research and standardization communities have launched an initiative called Universal Multimedia Access (UMA).[1,2] However, UMA tools and specifications that have so far emerged concentrate mostly on constraints imposed by terminals and networks along the multimedia delivery chain; users who consume the content are rarely considered. Pereira and Burnett[3] have presented some initial thoughts on how to incorporate user preferences and demands in today's multimedia infrastructures.

In practice, the main issues are as follows. On the one hand, end users want rich multimedia content available to them anytime, anywhere on nearly any kind of device. Because each device accessing the content has different capabilities—for example, display resolution or color depth—the content must be adapted according to these capabilities. On the other hand, content providers aspire to offer multimedia-based information of the best possible quality with respect to the consumer's context, without neglecting economic principles. The many problems providers face include supporting the various coding formats (without wasting disk space) and providing appropriate adaptation modules.

In recent years, researchers have developed a plethora of technologies and standards to address some of these issues. However, the big picture of how these different technologies and standards fit together is missing. Thus, the Moving Picture Experts Group (MPEG) decided to standardize the MPEG-21 Multimedia Framework[4] with the ultimate goal to support users during the exchange, access, consumption, trade, or other manipulation of so-called Digital Items in an efficient, transparent, and interoperable way.

Editor's Note:

The drive for innovation in multimedia technology often results in a diversity of ideas and approaches. However, this diversity is introducing new challenges for accessing multimedia content anytime, anywhere, any way.

The richness in multimedia content and increasing heterogeneity of networks and user devices is making interoperable multimedia communications difficult. Intelligent solutions are needed to enable multimedia content access under a wide range of delivery conditions and usage environments. The MPEG-21 standard promises to fill this need by standardizing descriptors for multimedia content access and allowing standards-compatible technologies to be used for adapting multimedia content.

In this article, Christian Timmerer and Hermann Hellwagner tell us how MPEG-21 seeks to achieve interoperable multimedia communication across networks and devices. They describe how MPEG-21 addresses device and format coding independence by standardizing descriptors of usage environment and bit-stream syntax. Through detailed examples involving streaming of audio–video resources and adapting of images according to terminal capabilities, the article illustrates how MPEG-21 solves the challenging and important problem of universal multimedia access.

—John R. Smith

Device and coding-format-independent multimedia content adaptation

Standardization committees such as MPEG, the Internet Engineering Task Force (IETF), and the World Wide Web Consortium (W3C) are exploring device and coding format independence issues. Here, however, we focus on how we can use the tools specified within MPEG-21 for interoperable multimedia communication.

1070-986X/05/$20.00 © 2005 IEEE

Published by the IEEE Computer Society

Device independence

Device independence, usually referred to as the ability to play (multimedia) content independently of the presentation device, requires an interoperable description format of the device's capabilities. The usage environment description (UED), part of the MPEG-21 Digital Item Adaptation (DIA) specification,[5-7] specifies how to describe such devices in terms of their codec capabilities, I/O capabilities, and device properties; furthermore, the UED defines description formats for the networks through which the content is accessed. The UED also provides a means for describing user characteristics and preferences—such as the user's current geographical location—and for describing the natural environment, such as illumination of the user's room. By facilitating descriptions of the user's environment, where the multimedia content is likely to be consumed, the UED contributes to maximizing the user's overall experience.

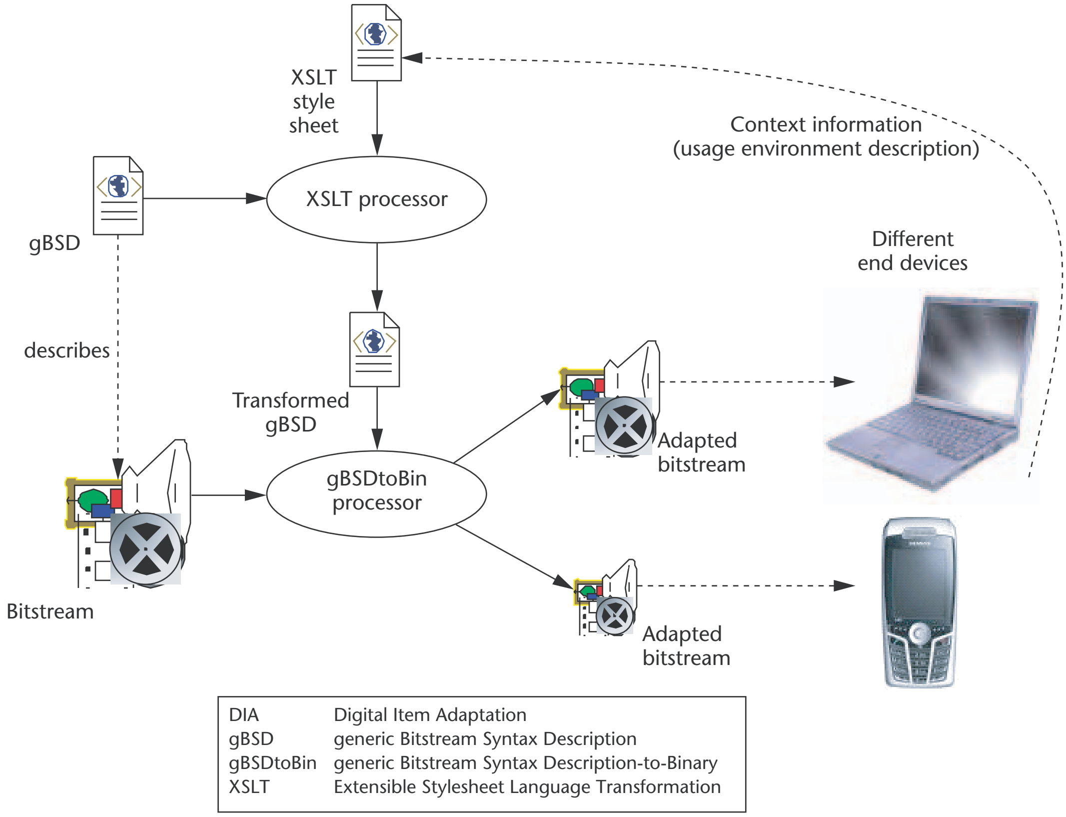

Figure 1. Multimedia publishing using MPEG-21 DIA according to different usage contexts.

Coding-format independence

So far, we've described only one aspect of interoperable multimedia communication: the consumer side of a media delivery and adaptation chain. To provide multimedia-based adaptation services, we must also consider the provider side's needs. The MPEG-21 Digital Item Declaration (DID)[8] provides a generic container format for associating metadata with multimedia content. DIDs containing references to the actual multimedia resources, and their associated metadata, supply the provider-side input to an adaptation engine. Here, we'll focus on the metadata in the DIDs that enables coding-format-independent multimedia adaptation.

To deal with the diversity of existing scalable coding formats—MPEG-4 or JPEG2000, say—it's desirable to have a generic approach for adapting media bitstreams. Equally important, such a generic approach lets us address emerging formats, such as MPEG-21 Scalable Video Coding. Within DIA, bitstream syntax description (BSD)[9] lets us describe the bitstream's high-level structure—for example, stream organization by frames, layers, or packets—using the Extensible Markup Language (XML). In particular, the generic BSD (gBSD), as described elsewhere,[10] allows us to perform coding-format-independent bitstream adaptations. According to this concept, we modify the gBSD of a media bitstream, for example, by means of an Extensible Stylesheet Language Transformation (XSLT) style sheet. Such modifications let us discard gBSD portions corresponding to specific frame types, or update certain layer information. The transformed gBSD is then input to a universal adaptation module that generates an adapted bitstream based on the gBSD information. With this approach, we can use a single generic adaptation module rather than a specific module for each existing or future coding format. Figure 1 shows the multimedia publishing framework using the MPEG-21 Digital Item Adaptation's gBSD and XSLT to adapt multimedia resources according to different user environments.

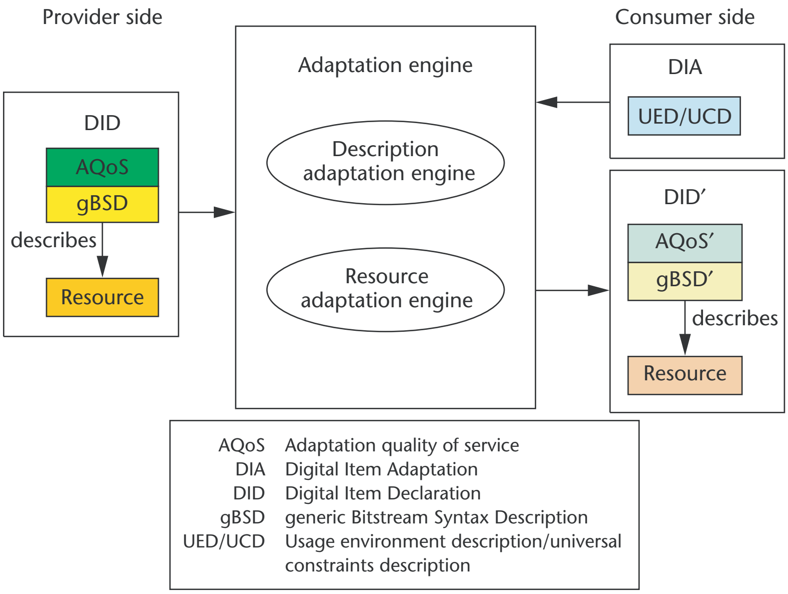

Figure 2. High-level architecture for device and coding-format-independent multimedia content adaptation using MPEG technologies.

The gBSD describing the high-level structure of the bitstream in Figure 1 is transformed according to the context information that different end devices provide. In particular, this context information sets the parameters of the XSLT style sheet responsible for transforming the gBSD. The transformed gBSD and the source bitstream are subject to the generic-bitstream-syntax-description-to-binary (gBSDtoBin) processor that performs the actual bitstream adaptation. This adaptation module parses and interprets the transformed gBSD, which describes the adapted bitstream by referencing the corresponding portions from the source bitstream. The gBSDtoBin processor simply copies these portions from the source bitstream to the target (that is, adapted) bitstream transmitted to the end device or user. Parameters that have been possibly updated during the transformation process are encoded in the adapted bitstream using the type information from the transformed gBSD. As such, the gBSD-based adaptation framework enables coding-format-independent adaptations.

However, the gBSD's transformation requires the optimum adaptation parameters with respect to the UED, taking into account quality of service (QoS) information. Two tools within DIA let us meet the above requirement, namely the AdaptationQoS (AQoS) tool and universal constraints description (UCD) tool, as Mukherjee et al. explain.[11] AQoS specifies the relationship between, for example, device constraints, feasible adaptation operations satisfying these constraints, and associated utilities (or qualities) of the multimedia content. On the other hand, the UCD lets us specify further constraints on the user environment and the use of a Digital Item by means of limitation and optimization constraints. An example is the information that the window size of the application is smaller than the actual display resolution. For example, the UED might describe a 1,400 × 1,050-pixel resolution, and the UCD constrains this further by informing the adaptation engine that only 70 percent of this is available.

Therefore, metadata like gBSD, AQoS, and UCD associated with media resources are packaged into DIDs, which forms the provider-side input to an adaptation module that enables interoperable multimedia communication.

Adaptation architecture

Figure 2 shows our proposed high-level architecture for device and coding-format-independent multimedia content adaptation. The DID, input to an adaptation engine from the provider side, contains the AQoS, the gBSD, and the media resource. The UED, which might be further constrained by UCDs, is the input from the consumer side. The description adaptation engine performs adaptation decision-making by processing the AQoS description and the UED and UCD. This engine then makes an optimal decision for the subsequent resource adaptation.

Within the resource adaptation engine, XSLT style sheets—using the parameters obtained from the decision-making process—transform the gBSD. Thereafter, the adaptation engine uses the transformed gBSD to generate the adapted media resource via the generic adaptation module, as specified within the MPEG-21 DIA. Finally, the adaptation engine transmits the adapted resource to the user where it's consumed during a multimedia experience. To allow for further adaptation steps, we package the updated gBSD and AQoS descriptions into a DID together with the adapted resource.

Because of the open interfaces, we can use this kind of adaptation engine anywhere within the multimedia delivery chain. Traditionally, it's on the server—or on an adaptation service—within intermediary network nodes.

Usage scenarios

Next, we present two usage scenarios that may benefit from such adaptation services:

- streaming of audio–video (AV) resources and

- adapting images according to terminal capabilities.

Streaming of AV resources

In today's ever-more-popular streaming applications, AV resources are usually preconfigured for various types of network conditions. In some cases, they're adapted within the rendering device according to its capabilities, such as those concerning color depth or spatial resolution. However, to follow this approach raises two issues:

- Storage capacity is wasted because of the maintenance of multiple copies of the same content for different classes of networks. Additionally, users are forced to configure their connectivity information to obtain an optimal user experience.
- Network bandwidth may still be wasted in cases where a high-quality bitstream is delivered to a device with limited color capabilities or spatial resolution. Also, central processing unit power is wasted in processing such a high-quality stream, which in turn consumes battery power.

Overcoming these problems requires that we store a single high-quality AV resource, adapted on the fly according to the capabilities of the requesting device and the characteristics of the networks traversed. Furthermore, with tools—as specified within the MPEG-21 DIA—we can consider user preferences as well as the natural environment, such as noise level or illumination.

Because the MPEG-21 DIA enables coding-format independence, we can describe the structure of the AV resources using gBSD and formulate the possible adaptation operations using AQoS. In some cases, we need to further constrain the user environment, which can be accomplished by a UCD, specifying—for instance—the application's window size.

Figure 3 illustrates multimedia adaptation as a service in intermediary network nodes. The consumer requests a resource from the provider, describing the usage context in the request. The provider redirects this request together with the resource (packaged into a DID) to an MPEG-21 DIA-compliant adaptation service, which performs the adaptation and forwards the adapted resource to the consumer. Note that, according to MPEG-21, Digital Items are exchanged between the involved parties (a provider and a consumer). A Digital Item comprises the request and the gBSD, AQoS, UED, and UCD information, and the adaptation service provides the user (consumer) with a Digital Item containing metadata and resources.

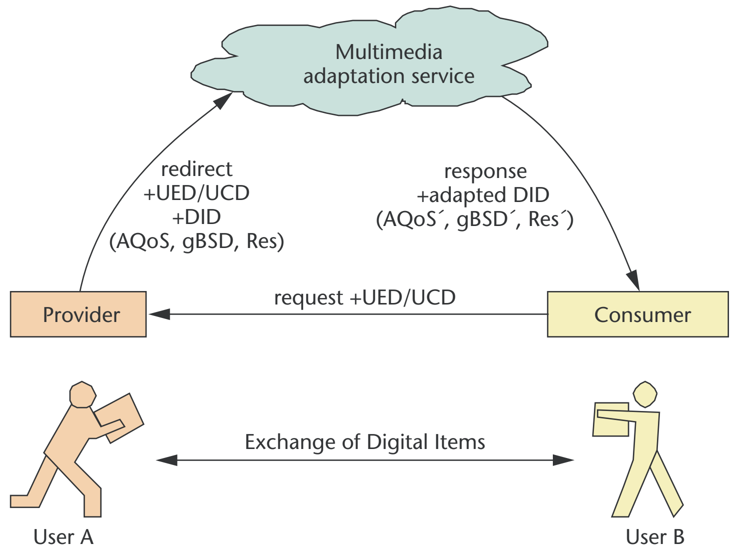

Figure 3. Multimedia adaptation as a service in intermediary nodes.

Adapting images according to terminal capabilities

A growing variety of mobile terminals will access multimedia resources of different formats. Especially in the Web community, the usage and richness of such resources will continue to increase.

Consider images as an example: They're already an integral part of almost every Web page, including private ones, now that digital cameras are available at reasonable prices. Providing an image tailored for each access device, network, and codec would multiply the images' storage and offline transcoding requirements. This solution is clearly neither economically feasible nor "future proof" because future image formats can't be considered. Therefore, we propose to exploit scalable image formats such as JPEG2000 together with the MPEG-21 tools introduced earlier. Our adaptation methodology—as Figures 1, 2, and 3 show—results in several versions of the same high-quality image satisfying the different devices' capabilities, which Figure 4 depicts.

In Figure 4, different mobile devices access the same high-quality JPEG2000-encoded image and indicate their display capabilities by means of MPEG-21 DIA descriptions. The optimal adaptation parameters for the images are provided by the associated AQoS descriptions—that is, the adaptation decision-taking engine (ADTE) determines

Figure 4. Different mobile devices access the same high-quality JPEG2000-encoded image resource, using MPEG-21 concepts.

how many decomposition levels and color components need to be dropped, given the display's spatial resolution and color properties. The actual bitstream adaptation is performed by means of the gBSD-based adaptation framework. (For additional information, see the "Further Reading" sidebar.)

Communication protocols

We've thus far focused on device and coding-format-independent bitstream adaptation. In practice, the information assets need to be communicated among the participants in the MPEG-21 multimedia framework. MPEG-21 itself doesn't provide any standardized protocols for this task, but other standardization bodies like the IETF or W3C do. Some protocols that can be used for this purpose follow.

Real-time Transport Protocol (RTP)

This protocol is especially suitable for real-time streaming of AV resources. Associated metadata could be transported together with the resource within the custom header in each RTP packet. Alternatively, a separate RTP stream could be set up for the metadata, but this increases synchronization efforts at the involved nodes (see RFC3550 for further details, http://www.ietf.org/rfc/rfc3550.txt).

Real-Time Streaming Protocol (RTSP)

RTSP provides primitives (Setup, Play, Pause, Describe, and so on) for negotiating AV resources over IP-based networks. RTSP often uses the Session Description Protocol (SDP, see RFC2327 for further details, http://www.ietf.org/rfc/rfc2327.txt) for providing information relevant to the current session, such as the current frame rate of a video or the network bandwidth. This is, however, proprietary and protocol-specific and doesn't use MPEG-21 DIA UEDs, which are designed to be protocol independent based on XML. To close this gap, the next generation of the SDP (SDPng), which is being developed, is capable of transporting XML and, therefore, UEDs during the negotiation phase (see RFC2326 for further details, http://www.ietf.org/rfc/rfc2326.txt).

Session Initiation Protocol (SIP)

SIP is a text-based protocol for initiating interactive communication sessions between users. Such sessions include voice, video, chat, interactive games, and virtual reality. In this sense, it's similar to SDP and could benefit from the SDPng development as well (see RFC3261 for further details, http://www.ietf.org/rfc/rfc3261.txt).

Hypertext Transfer Protocol (HTTP)

The W3C's protocol for the Web targets more or less nonstreaming resources because of its TCP-based architecture. However, extensions to HTTP also allow the inclusion of capabilities descriptions, which could be used to include UEDs within the resource request. Metadata associated with the requesting resource can be handled like the resource itself (see RFC2616 for further details, http://www.ietf.org/rfc/rfc2616.txt).

Further Reading

For more information about MPEG-21, including overviews, FAQs, and working documents, see http://www.chiariglione.org/mpeg.

Additional MPEG-21 DIA material can be found at http://mpeg-21.itec.uni-klu.ac.at/.

Future work

The MPEG-21 DIA provides tools for guaranteeing interoperability for media adaptation purposes in the entire multimedia delivery chain. The main focus is on enabling device and coding-format-independent adaptation engines that various industries and businesses can construct. Therefore, the interfaces to such adaptation engines have been standardized, and it's up to the industry to compete for implementations compliant with the MPEG-21 DIA.

However, the MPEG-21 DIA doesn't standardize any transport or messaging mechanisms enabling the communication of such descriptions. This task needs to be accomplished one level below the application layer on which MPEG-21 operates—that is, by means of communication protocols as we've outlined. It's conceivable that bindings to the various protocols need to be defined within other standardization bodies. Furthermore, the transport of media and associated metadata raises several synchronization issues that need to be solved; in doing so, researchers need to clearly keep efficiency in mind. In particular, metadata's significant size—in terms of file size—and XML's verbosity need to be reduced. An alternative (binary) serialization of XML data is one step toward this goal, and MPEG-7 Systems[12] seems to be a promising candidate. On the other hand, the W3C's XML Binary Characterization Working Group (http://w3.org/XML/Binary/) has recently started investigating this issue but hasn't produced a recommendation so far. **MM**

Acronyms

ADTE	Adaptation decision-taking engine
AQoS	Adaptation QoS
BSD	Bitstream syntax description
DIA	Digital Item Adaptation
DID	Digital Item Declaration
gBSD	generic Bitstream Syntax Description
gBSDtoBin	generic-Bitstream-Syntax-Description-to-Binary
IETF	Internet Engineering Task Force
MPEG	Moving Picture Experts Group
QoS	Quality of service
SDP	Session Description Protocol
SDPng	Session Description Protocol, next generation
UCD	Universal constraints description
UED	Usage environment description
UMA	Universal Multimedia Access
W3C	World Wide Web Consortium
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformation

Acknowledgments

The authors wish to thank Jörg Heuer, Gabriel Panis, and Andreas Hutter from Siemens AG; Eric Delfosse from IMEC; Debargha Mukherjee from HP Labs; Sylvain Devillers from France Telecom R&D; and Harald Kosch from Klagenfurt University for contributing various parts to the standards we've discussed.

References

1. J. Smith, R. Mohan, and C.-S. Lee, "Scalable Multimedia Delivery for Pervasive Computing," *Proc. ACM Multimedia*, ACM Press, 1999, pp. 131-140.
2. A. Vetro, C. Christopoulos, and T. Ebrahami, eds., *IEEE Signal Processing Magazine*, special issue on Universal Multimedia Access, vol. 20, no. 2, 2003.
3. F. Pereira and I. Burnett, "Universal Multimedia Experiences for Tomorrow," *IEEE Signal Processing Magazine*, vol. 20, no. 2, 2003, pp. 63-73.
4. I. Burnett et al., "MPEG-21: Goals and Achievements," *IEEE MultiMedia*, vol. 10, no. 6, Oct.–Dec. 2003, pp. 60-70.
5. A. Vetro and C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities," *IEEE Trans. Multimedia*, vol. 7, no. 2, April 2005.
6. A. Vetro, C. Timmerer, and S. Devillers, "Digital Item Adaptation," *The MPEG-21 Book*, R. Van de Walle et al., eds., John Wiley & Sons, 2005 (to appear).
7. *Information Technology—Multimedia Framework (MPEG-21)—Part 7: Digital Item Adaptation*, ISO/IEC 21000-7, 2004. For information, see http://www.iso.org/; also http://www.chiariglione.org/mpeg.
8. F. De Keukelaere and R. Van de Walle, "Digital Item Declaration and Identification," *The MPEG-21 Book*, R. Van de Walle et al., eds., John Wiley & Sons, 2005 (to appear).
9. G. Panis et al., "Bitstream Syntax Description: A Tool for Multimedia Resource Adaptation within MPEG-21," *EURASIP Signal Processing: Image Communication J.*, vol. 18, no. 8, 2003, pp. 721-747.
10. C. Timmerer et al., "Coding Format Independent Multimedia Content Adaptation Using XML," *Proc. SPIE Int'l Symp.* (ITCom 03), vol. 5242, SPIE, 2003, pp. 92-103.
11. D. Mukherjee et al., "Terminal and Network Quality of Service," *IEEE Trans. Multimedia*, vol. 7, no. 2, April 2005.
12. J. Heuer, C. Thienot, and M. Wollborn, "Binary Format," *Introduction to MPEG-7: Multimedia Content Description Language*, B.S. Manjunath, P. Salembier, and T. Sikora, eds., John Wiley & Sons, 2002.

Readers may contact Christian Timmerer and Hermann Hellwagner at {christian.timmerer, hermann.hellwagner}@itec.uni-klu.ac.at.

Contact Standards editor John R. Smith at jsmith@us.ibm.com.