Towards MPEG-21-based Cross-layer Multimedia Content Adaptation

Ingo Kofler, Christian Timmerer,
Hermann Hellwagner
Department of Information Technology,
Klagenfurt University, Austria
<firstname.lastname>@itec.uni-klu.ac.at

Toufik Ahmed

CNRS-LaBRI Lab,
University of Bordeaux-1, France
tad@labri.fr

Abstract

Cross-layer designs are becoming more and more attractive within the multimedia community since multiple-play services pave their way towards consumer markets enabling mobility in various aspects. However, cross-layer designs so far have mainly focused on performance issues and do not provide much support in terms of interoperability which is a requirement for services envisaged as part of the Fixed-Mobile Service Convergence (FMSC) initiative. This paper presents a first attempt towards increasing the interoperability of cross-layer designs by adopting an open standard – MPEG-21 Digital Item Adaptation – for describing the functional dependencies across network layers. In this paper a three-step approach for multimedia content adaptation is presented that introduces an MPEG-21-based cross-layer architecture.

1. Introduction

Fixed-Mobile Service Convergence (FMSC) is gaining momentum enabling end user device mobility, service mobility, and personal mobility. The end user mobility allows the user to make use of his devices independent of his location. The service mobility provides means for seamless service delivery independently of the end user's device, access network, and location. Finally, the personal mobility guarantees ubiquity across different network domains. In this regard, multimedia services which have been traditionally designed for fixed-wired networks expand into the mobile environment and are becoming an integral part of FMSC. However, these multimedia services have tough requirements regarding bandwidth, delay, jitter, and packet loss which are not very well supported in wireless and mobile environments.

The Internet Protocol suite with its well-defined layers and interfaces is one rigid obstacle that hampers the deployment of multimedia services in wireless networks. Furthermore, the wide deployment of wireless networks created the demand for enhancing the performance of multimedia applications over these wireless links. Cross-layer designs are currently an active research topic aiming at increasing the Quality of Service/Experience (QoS/QoE) by performing coordinated actions across the network layers and, thus, violating the protocol hierarchy and isolation model. As a result, a variety of different approaches emerged in the last years [1][2]. In the majority of these approaches, the cross-layer interactions take place in either a bottom-up or a top-down fashion, where upper layers influence lower layers or vice versa. More recent efforts are undertaken in the form of tackling the problem by jointly optimizing parameters at the different layers [3]. However, independently of the ways the different cross-layer designs perform, they all share the common property of compromising interoperability in favor of performance.

In this paper, we present an approach which aims to increase the degree of interoperability. Our approach is based on utilizing description tools and algorithms standardized within MPEG-21 Digital Item Adaptation (DIA) [2] to increase user QoS/QoE. This approach comprises three steps which are outlined in the following and detailed in the remainder of this paper. The first step identifies possible cross-layer interactions in terms of dependencies among multimedia characteristics and usage environment conditions across different network protocol layers. Note that in our approach, we currently focus on scalable video content such as MPEG-4 Scalable Video Coding (SVC). The result of this first step is referred to as the Cross-Layer Model (XLM) which is then – in a second step – formulated by utilizing interoperable description tools as standardized within MPEG-21 DIA. The third step comprises finding an

Acknowledgment: This work is supported by the European Commission in the context of the ENTHRONE project (IST-1-507637). Further information is available at http://www.ist-enthrone.org

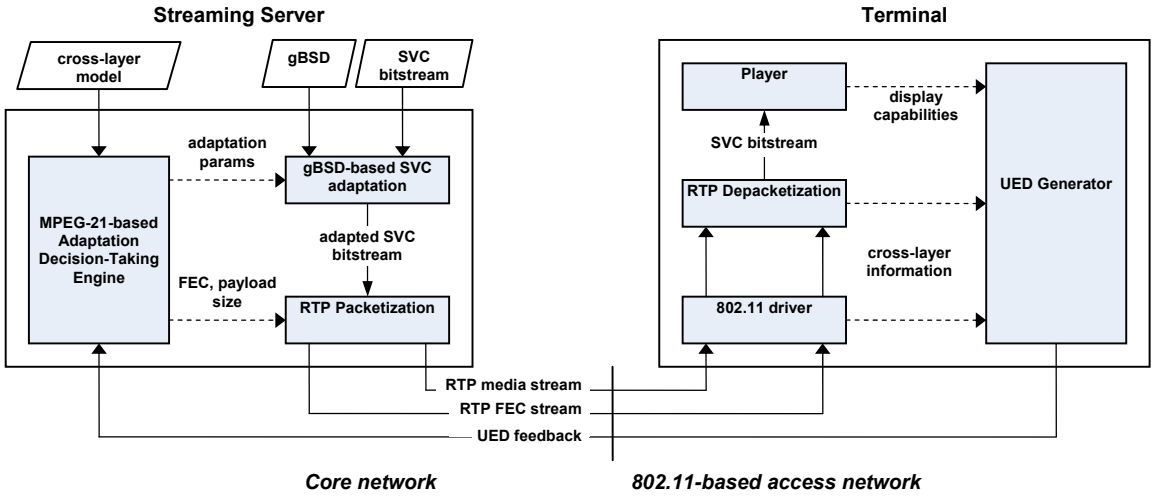

Figure 1. Architecture of the MPEG-21-based cross-layer approach

optimal solution for the optimization problem at hand by means of a generic metadata-driven Adaptation Decision-Taking Engine (ADTE).

2. MPEG-21-based Cross-layer Approach

In our approach we focus on the streaming of scalable (video) content to a wireless terminal. The content is delivered from a streaming server which is directly connected to a fixed-wired core network. At the content consumer's side of the transmission chain, a wireless terminal is connected to the core network through an access network. An 802.11g-based wireless LAN is used as the access network. The core network is assumed to provide QoS mechanisms to ensure the transmission of real-time multimedia traffic according to pre-negotiated network parameters (e.g., delay, jitter, packet loss) between providers (i.e., Service Level Agreements). Additionally, an admission control algorithm is employed to reject video streams that would exceed the actual capacities of the core network. Therefore, in this QoS-enabled core network the QoS parameters are statistically engineered. On the access network, the 802.11 link does not provide any of the abovementioned guarantees since packet loss is unavoidable due to the nature of wireless communication such as interference, channel fading, and signal attenuation. These effects lead to unreliable networking conditions without any QoS guarantees concerning the available bandwidth and packet loss. In order to provide a smooth video playback at the wireless terminal, the video stream has to be adapted dynamically according to the changing conditions of the wireless access network using a cross-layer approach. In this paper, the MPEG-21-based description tools are adopted for both steering and performing the cross-layer adaptation of the video stream. The architecture of both the streaming server and the wireless terminal is depicted in Figure 1.

The adaptation of the SVC video bitstream is performed at the streaming server using a generic MPEG-21-based adaptation engine [3]. The normative generic Bitstream Syntax Description (gBSD) tool is used to describe the frames and their corresponding Network Adaptation Layer (NAL) units of the bitstream. The description is based on XML and describes the offset and the length of each NAL unit and whether they belong to the temporal, spatial, and signal-to-noise ratio (SNR) scalability layer. The actual adaptation is performed in two steps. The first step is performed in the metadata domain. In fact, the gBSD that describes the bitstream is adapted by removing the parts that describe the NAL units of the bitstream that have to be removed, e.g., NAL units belonging to a certain temporal layer. The adaptation of the metadata is performed by an XSLT processor that transforms the initial gBSD based on a parameterized style sheet. The second step of the generic adaptation is the modification of the actual bitstream according to the transformed gBSD. This step is performed within the normative gBSDtoBin process. The input to this process is both the bitstream and the adapted gBSD.

The output is an adapted bitstream that reflects the changes made in the metadata description. It should be noted that in the case of video streaming the adaptation is not performed on the whole bitstream at once but on a per-picture basis.

The adapted SVC bitstream is then streamed using the RTP protocol. The NAL units are packetized according to [4]. Depending on the actual size of the NAL units the packetizer dynamically switches between Single-Time or Multi-Time Aggregation Packets and Fragmentation Unit packets, leading to an efficient packetization process. The packetizer can be dynamically configured to generate packets with a given maximum payload size. The RTP packetizer can optionally generate a second RTP stream that carries Forward Error Correction (FEC) packets according to the payload format for generic FEC [5]. The FEC stream can be used to recover from lost media RTP packets by reconstructing them based on the other original media and FEC packets. The packetizer calculates the FEC packets using a linear block code. The amount of FEC packets that are generated can be configured dynamically by specifying the (n, k) parameters for the block code. This means that (n-k) FEC packets will be used to protect a block of k media packets. This basically provides a resiliency against a maximum packet loss rate of p=(n-k)/n when considering that also FEC packets are affected by loss.

3. Cross-layer Interactions

The main motivation of our approach is to control the video streaming at the application layer by utilizing the dynamic MPEG-21-based adaptation and flexible packetization according to the conditions of the wireless 802.11 link. Basically, we will focus on three different cross-layer interactions.

At the physical layer the 802.11g standard offers a variety of different modulation and coding rates that result in different physical rates. The offered physical rates range from 1 Mbps (BPSK, coding rate 1/11) to 54 Mbps (QAM-64, coding rate 3/4) and represent a trade-off between robustness and achievable link capacity. 802.11 network interfaces are exploiting this trade-off by dynamically adapting the physical rate according to the link quality. This mechanism is referred to as adaptive rate selection (or rate control algorithm) and is not normative. However, the fact that the capacity of the link is time-varying can be exploited by explicitly signaling the physical rate to the application layer leading to a significant improvement of the video quality at the terminal [5]. In our architecture, the explicit feedback is used to steer the adaptation of the video to prevent an excessive load at the wireless link.

At the data-link layer, the achievable throughput over 802.11 networks highly depends on the payload size of the packet. As a consequence of the underlying MAC scheme, the throughput is very low at small packet sizes. Additionally, the efficiency is further reduced by the fixed size of the higher protocol headers that lead to a significant payload overhead for small payload sizes. For that reason a common approach is to maximize the payload size and use the path Maximum Transmission Unit (MTU) as an upper limit to avoid fragmentation at the IP level. However, this might lead to a suboptimal performance at wireless networks because of the comparatively high Bit Error Rates (BER). Since the packet error rate depends on both the BER and the size of the packet, the probability of having an uncorrectable packet error at the receiver is higher for larger packets than for smaller ones. This behavior can be utilized to maximize the throughput by selecting a payload size that is optimal for the actual link conditions [6]. In our approach, this knowledge will be exploited by providing the optimum payload size to the RTP packetizer in order to ensure that the majority of the generated RTP packets have a throughput-optimized size.

At the transport layer, packet loss of RTP media packets is addressed. In comparison with wired links, the wireless transmission is less reliable and packets can be lost during transmission over the air due to several reasons. However, 802.11 network devices adopt retransmission schemes at the data-link layer only up to a certain limit of retransmission attempts. If the upper limit of the retransmission attempts is reached the data-link frame is discarded. Normally, these losses are handled by transport layer protocols such as TCP that offer reliable transmission based on end-to-end retransmission. But in the case of video streaming, these retransmissions are not desirable since they introduce additional delay and jitter. Therefore, a separate application-level FEC stream is needed to allow the reconstruction of lost packets at the terminal. The amount of FEC that is required for enabling a smooth playback at the receiver depends on the actual condition of the wireless link. The code rate of the FEC packetizer will be configured dynamically based on the feedback from the data-link and physical layers. The received signal strength and the percentage of non-decodable frames are used as indicator for the actual quality of the link.

4. The Cross-layer Model

The control logic for the cross-layer adaptation is represented by the cross-layer model (XLM). The cross-layer model basically describes how the adaptation and packetization parameters are determined with respect to the limitations of the wireless network and the terminal. The XLM is represented as a mathematical optimization problem. For that reason, both the parameters (temporal, spatial, and SNR layers, payload size and FEC code rate), the resulting content properties (e.g., video bitrate, PSNR value), and the usage context (e.g., physical rate, signal strength) are modeled as mathematical variables. While the values for the variables that are representing adaptation parameters can be regarded to be chosen arbitrarily from a given set of possible values (=domain), the variables representing the usage context are bounded to the measured values.

Among the variables, there exist functional relationships that are also expressed within the XLM. This means that a value of a variable may be determined by a mathematical function that uses one or many other variables as arguments. For example, the video bitrate can be seen as function of the three variables that are representing the temporal, spatial, and SNR layers of the scalable bitstream. Since the video stream offers only a finite number of adaptation possibilities, the function can be described by explicitly listing all possible layer combinations and the resulting bitrate. In contrast to that, some functional dependencies have to be modeled using continuous functions. For example, the required network bandwidth for a video bitstream with a given bitrate is approximated considering the amount of FEC that is added and the IP/UDP/RTP protocol overhead that actually depends on the selected payload size.

Although the different adaptation dimensions offer a large number of adaptation possibilities, not all of them might be applicable. Some combinations of adaptation parameters might lead to a video bitrate that might overload the access network and should be avoided. To ensure that only feasible adaptations are considered, the XLM includes a variety of constraints, e.g., one constraint specifies that the video bitrate has to be at least lower than the actual physical downstream rate of the 802.11g link. This restricts the set of feasible adaptation possibilities and prevents from steering the adaptation in a way that would cause an excessive load at the wireless link. Other constraints are forcing the resolution of the video to not exceed the terminal's capability, avoiding situations that high definition content is streamed to handheld devices with a low-resolution display. Although constraints limit the set of feasible adaptations, there might be more than one feasible adaptation. The final selection of the actual adaptation that is enforced is steered by using an objective function. The XLM for the 802.11g access network includes one objective function that aims at maximizing the resulting quality of the video in terms of PSNR. The optimum adaptation parameters regarding the XLM are determined by finding appropriate values for the variables that do not violate any of the constraints and maximize the PSNR. The computational cost of determining the optimum parameters depends on the number of layers that are offered by the SVC bitstream. The discrete number of layers results in a finite number of adaptation possibilities that are in an order of some tens to few hundred possibilities. An evaluation of suitable optimization algorithms showed that an improved version of an exhaustive search in the parameter space is efficient enough for such a limited number of adaptation possibilities [10].

5. MPEG-21 Support for Adaptation

The adaptation and packetization at the streaming server is steered by the adaptation decision-taking process, which is responsible to determine the optimum adaptation parameters. These parameters are based on the adaptation capabilities of the video stream and the actual usage context including the network conditions. The MPEG-21 DIA standard defines three different tools to enable a generic adaptation decision-taking process.

The Usage Environment Description (UED) tool is used to describe network and terminal capabilities, preferences and impairments of the content consumer, and the natural environment in which the multimedia content is being consumed. For example, the available codecs at the terminal, the type of device (e.g., PC, PDA), display and audio playback capabilities can be signaled.

The available adaptation parameters, their effects on the content's properties and quality can be expressed using the AdaptationQoS tool. It provides means for declaring parameters and properties as kind of mathematical variables which are referred to as IOPins within the MPEG-21 terminology. The impact of certain adaptation parameters on the content's properties are expressed by functional dependencies between the IOPins which are called modules within MPEG-21 DIA. Three different types of modules are available including a look-up table mechanism and the

representation of the dependency as a function in postfix notation (stack function).

The third tool denoted as Universal Constraint Description (UCD) tool is used to specify constraints on the variables and to define objective functions which should be maximized or minimized (e.g., maximize the frame rate).

Based on the abovementioned three types of descriptions, a mathematical optimization problem can be derived [7]. The optimization problem can be solved by finding appropriate values for the variables (=IOPins) that do not violate the constraints and are optimal concerning the objective function(s). These values are then used as parameters for the actual adaptation. The advantage of this approach is that the actual control logic for the adaptation is defined via metadata while the software component that interprets the metadata – the Adaptation Decision Taking-Engine (ADTE) – remains generic.

AdaptationQoS and UCD tools are used to implement the XLM. The mathematical optimization problem can be transferred into MPEG-21-based XML descriptions as follows. IOPins are used to describe the variables and their domains, while stack functions and look-up tables express the functional dependencies among them. The constraints of the XLM are represented as limitation constraints, i.e., stack functions that evaluate to true or false. The objective function that aims to maximize the PSNR is expressed within a UCD by a maximization constraint.

The MPEG-21-based ADTE is located at the streaming server where it passes the adaptation decisions to the generic adaptation engine (adaptation parameters) and to the RTP packetizer (FEC and payload size). The UED that is required for the decision-taking is generated by the UED Generator at the wireless terminal and is transmitted to the streaming server in regular intervals by using a reliable transport protocol. The UED reflects the actual usage environment context at different layers according to the layered network protocol stack. At the physical layer, the actual downstream physical rate (from the access point to the wireless terminal), the number of non-decodable frames, and the received signal strength are determined and signaled by the UED. Furthermore, the information about the percentage of lost RTP packets and the jitter which is determined on the application layer (i.e., using RTCP feedback) is integrated into this description. In addition to the network-related information, the display capabilities (resolution and maximum frame rate) of the terminal are described to allow an optimum adaptation according to the capabilities of the terminal. It should be mentioned that the UED tool as specified within the standard is not intended to describe features and properties specific to a certain network technology or protocol. Therefore, we had to extend the initial XML schema of the UED tool to support 802.11 and RTP/RTCP specific properties.

6. Conclusions and Future Work

In this paper we presented an MPEG-21-based approach for cross-layer adaptation that leverages the usage of MPEG-21 metadata for both the adaptation of scalable video content and the adaptation decision-taking. The adaptation of the video is based on the generic Bitstream Syntax Description (gBSD) tool that allows a coding-format independent adaptation of scalable bitstreams. The adaptation and packetization is steered by an MPEG-21-based Adaptation Decision-Taking Engine, which is a generic component that operates on MPEG-21 metadata. The optimum parameters for the adaptation and packetization are determined based on a cross-layer model (XLM) along with a description of the actual usage context including the actual state of the wireless link.

The basic idea of our proposal is to improve the streaming of the SVC video content over the 802.11g link by dynamically choosing a throughput-optimized payload size, an optimum forward error correction at the RTP level and by adapting the content according to the physical rate of the wireless link. Information about the condition of the wireless link and statistics of the RTP stream are collected at different layers and are explicitly signaled to the application layer using the MPEG-21 Usage Environment Description. Additionally, terminal capabilities are considered for enabling an optimum adaptation. So far we defined the basic architecture and developed an initial cross-layer model. Based on that work, we want to determine a complete cross-layer model including all functional dependencies and investigate if and to which extent the three envisaged cross-layer interactions lead to a quantifiable as well as perceivable improvement of the video quality at the wireless terminal.

7. References

[1] M. van der Schaar and N. Sai Shankar, “Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms”, *IEEE Wireless Communications*, vol. 3, no. 4, pp. 50–58, Aug. 2005.

[2] V. Kawadia and P. Kumar, “A Cautionary Perspective on Cross Layer Design”, *IEEE Wireless Communications*, vol. 12, no. 1, pp. 3-11.

[3] L. Choi, W. Kellerer, and E. Steinbach, “On Cross-Layer Design for Streaming Video Delivery in

Multiuser Wireless Environments", *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, pp. 1-10, 2006.

[4] A. Vetro, "MPEG-21 Digital Item Adaptation: Enabling Universal Multimedia Acces", *IEEE Multimedia*, vol. 11, no. 1, pp. 84-87, Jan-March 2004.

[5] J. Rosenberg and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction", RFC 2733, 1999.

[6] C. Timmerer, G. Panis, and E. Delfosse, "Piece-wise Multimedia Content Adaptation in Streaming and Constrained Environments", *Proceedings WIAMIS'05*, Montreux, Switzerland, April 2005.

[7] S. Wenger, et.al., "RTP Payload Format for SVC Video", Internet Draft draft-ietf-avt-rtp-svc-01.txt, March 2007.

[8] I. Djama and T. Ahmed, "MPEG-21 Cross-Layer QoS Adaptation for Mobile Service Delivery", *Proceedings AXMEDIS 2006*, pp. 215–222, Leeds, UK, Dec. 2006.

[9] S. Choudhury, J.D. Gibson, I. Sheriff, and E.M. Belding-Royer, "Effect of Payload Length Variation and Retransmissions on Multimedia in 802.11a WLANs", *Proceedings IWCMC'06*, pp. 377–382, Vancouver, Canada, July 2006.

[10] I. Kofler, C. Timmerer, H. Hellwagner, A. Hutter, and F. Sanahuja, "Efficient MPEG-21-based Adaptation Decision-Taking for Scalable Multimedia Content", *Proceedings MMCN'07*, pp. 65040J-1 - 65040J-8, San Jose, USA, January/February 2007

[11] Mukherjee, E. Delfosse, J.G. Kim, and Y. Wang, "Optimal adaptation decision-taking for terminal and network quality-of-service", *IEEE Trans. on Multimedia*, vol. 7, no. 3, pp. 452–462, June 2005.