

An Interoperable Streaming Framework for Scalable Video Coding Based on MPEG-21^{*}

Michael Eberhard^{*}, Luca Celetto[†], Christian Timmerer^{*}, Emanuele Quacchio[†],
and Hermann Hellwagner^{*}

^{*}ITEC, MMC – Klagenfurt University, Austria, <firstname.lastname>@itec.uni-klu.ac.at

[†]ST Microelectronics, Italy, <firstname.lastname>@st.com

Keywords: Scalable Video Coding, Adaptation, Video on Demand, Multicasting, MPEG-21 Digital Item Adaptation.

Abstract

In this paper, an interoperable streaming framework for the adaptation and streaming of scalable multimedia content is presented. The streaming framework's architecture for Video on Demand (VoD) as well as multicast streaming is presented. The VoD test-bed allows each client to set up a separate streaming session and receive the bitstreams in a quality tailored to the clients requirements, while the multicast test-bed provides a layered scalable multicast to all clients and every client needs to decide which layers to subscribe to. After the description of the architectures, a comparison in terms of performance of the MPEG-21 DIA metadata-based adaptation approach to an SVC-specific adaptation approach is presented. Furthermore, optimizations for both adaptation approaches are presented and the performance improvements due to the optimizations are discussed.

1 Introduction

The streaming of multimedia content over heterogeneous networks like the Internet is still a challenging research area, as the capabilities of the users' terminals vary strongly in terms of supported codecs, display resolution, processing power or energy consumption. Furthermore, different users have access to network connections with varying bandwidth capabilities and the network conditions of these connections are frequently changing. The usage of technologies like transcoding or stream switching does not offer in general the performance and flexibility to provide the multimedia resources to the users in a quality best suited for their terminals [1]. For such heterogeneous environments, scalable multimedia coding formats provide a good possibility to satisfy the requirements of the users, as the adaptation of such scalable bitstreams can be performed by simple removing or minor editing operations. The most promising scalable video coding standard today is the Scalable Video Coding (SVC) extension of the ISO/ITU-T's Advanced Video Coding (AVC) standard [2], which provides support for temporal, spatial and

signal-to-noise ratio (SNR) scalability while still maintaining the superior coding efficiency of AVC.

Although scalable multimedia coding formats can be easily adapted, this process is different for each of the various coding formats. To avoid the implementation of a specific adaptation engine for each coding format, interoperable adaptation engines – which perform the adaptation process independent of the underlying coding scheme – are required. Part 7 of the MPEG-21 standard, Digital Item Adaptation (DIA) [3], provides a solution for such an interoperable adaptation engine based on metadata.

In this paper, an interoperable framework for the streaming and adaptation of scalable multimedia contents is presented. The framework provides the possibility for Video on Demand (VoD) streaming as well as multicast streaming deployed for Internet Protocol Television (IPTV) applications. The architecture of both implementations is discussed in detail in Sections 2 and 3. To evaluate the performance of the generic MPEG-21 DIA metadata-based adaptation approach in comparison to an SVC-specific adaptation approach, our implementation was compared to the Bitstream Extractor of the JSVM reference software [4]. First results have shown that the main bottleneck of the metadata-based adaptation approach is the processing of the metadata. Thus, an alternative to the traditionally used Extensible Stylesheet Language for Transformations (XSLT) libraries is investigated. Furthermore, the reference software's Bitstream Extractor was optimized as well by utilizing length information in the Network Abstraction Unit (NAL) header. The results of the performance measurements for both adaptation approaches are presented in Section 4 and show that such optimizations can significantly improve the performance of the adaptation process.

2 An Interoperable Architecture for Video on Demand

The architecture presented in this section is illustrated in Figure 1 and provides the possibility to adapt scalable multimedia streams for VoD sessions. The implementation consists of the *MPEG-21 DIA-enabled VOD Server*, which performs the adaptation of the scalable bitstreams based on

^{*} This work is supported in part by the European Commission in the context of the P2P-Next project (FP7-ICT-216217). Further information is available at <http://www.p2p-next.org/>.

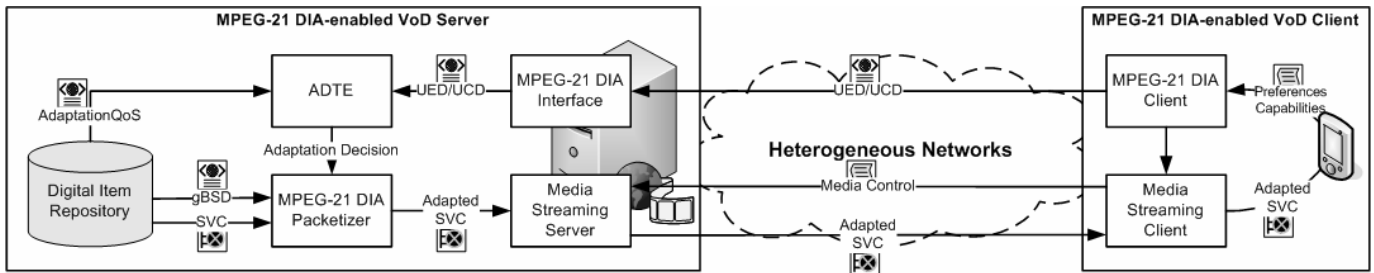


Figure 1: An Interoperable Architecture for Video on Demand.

MPEG-21 DIA metadata, and a number of *MPEG-21 DIA-enabled VoD Clients*, which receive the bitstreams in separate VoD sessions and in their preferred quality.

The streaming capabilities for the test-bed implementation are provided by the VLC media player and streaming server¹, which is used at the server as well as at the client side. The VLC has been extended by the MPEG-21 DIA tools, which are used for the adaptation of the bitstreams tailored to the capabilities and preferences of the clients. A streaming run course for the VoD test-bed is given in the following.

At the beginning of the streaming session, the client sends the preferences of the user as well as the capabilities of the user's terminal to the server. This context-related information is submitted to the server formatted as MPEG-21 DIA Usage Environment Description and as MPEG-21 DIA Universal Constraint Description [3]. The *MPEG-21 DIA Interface* receives the metadata at the server and passes the descriptions to the *Adaptation Decision-Taking Engine (ADTE)*. The *ADTE* subsequently takes the adaptation decision based on the context-related metadata received from the client and the *AdaptationQoS* description [3], which describes the adaptation capabilities of the content. The optimal adaptation parameters are found by matching the usage environment properties from the client with the available *SVC* layers described by the *AdaptationQoS* description. Furthermore, the *MPEG-21 DIA Interface* sets up the VoD streaming session, returning the corresponding Real Time Streaming Protocol (RTSP) URL to the client.

The actual adaptation of the *SVC* bitstreams is performed by the *MPEG-21 DIA Packetizer*. The packetizer receives the adaptation parameters from the *ADTE* and is called by the VLC for the processing of every access unit. The adaptation is performed utilizing the adaptation parameters and the generic *Bitstream Syntax Description (gBSD)* [3], which provides a high-level description of the structure of the bitstream. For more details of the adaptation process the reader is referred to [5]. After the adaptation process, the adapted access units are streamed to the client.

In addition to the initial sending of the context-related metadata from the client to the server, the test-bed allows the client to dynamically update their properties. If the user wants to update some properties, the *UED* and *UCD* are sent to the server, the adaptation decision is taken again and the updated parameters are provided to the packetizer. As the packetizer is called for the adaptation of each access unit separately, the update of the adaptation parameters is visible as soon as the next access unit is received at the client. Due to space

constraints, the reader is referred to [5] for more details about the test-bed.

3 An Interoperable Architecture for Multicast Streaming

The second test-bed implementation is illustrated in Figure 2 and provides the layers of an *SVC* bitstream in several parallel multicast Real-Time Transport Protocol (RTP) sessions. The multicast streaming approach aims to save bandwidth by providing the same layered stream to all clients, while each client can decide which layers to subscribe to. Furthermore, the clients can dynamically subscribe/unsubscribe one of the RTP session, in case the usage environment properties change.

The multicast architecture consists of one *MPEG-21 DIA-enabled Multicast Server* and a number of *MPEG-21 DIA-enabled Multicast Clients*. To enable the clients to subscribe to the multicast session, the information about the session is announced continuously by the *Media Streaming Server* in the *SDP (Session Description Protocol)*² format utilizing the *Session Announcement Protocol (SAP)*³. Furthermore, the server continuously performs a layered multicast of all the scalable layers in separate RTP sessions.

The *SAP/SDP Recipient* at the client receives the *SDP* announcements of all available bitstreams and saves them in the playlist of the available channels. If the user decides to subscribe to one of the *SVC* multicasts, the decision which layers need to be subscribed is taken. To take this adaptation decision, the client generates the *AdaptationQoS* description based on the properties of the *SVC* layers in the *SDP* description. The user preferences and the details about the capabilities of the user's terminal are already available at the client and are provided formatted as *UED/UCD* to the *ADTE*, together with the generated *AdaptationQoS* description. The *ADTE* takes the adaptation decision based on these metadata and passes the decision which layers need to be subscribed to the *Media Streaming Client*. The adaptation decision is subsequently mapped to the layers described by the *SDP* description and the corresponding layers are subscribed and received by the *Media Streaming Client*. Due to space constraint, the reader is referred to [5] for further information on the multicast streaming test-bed.

¹ <http://www.videolan.org/vlc/>

² <http://www.ietf.org/rfc/rfc2327.txt>

³ <http://www.ietf.org/rfc/rfc2974.txt>

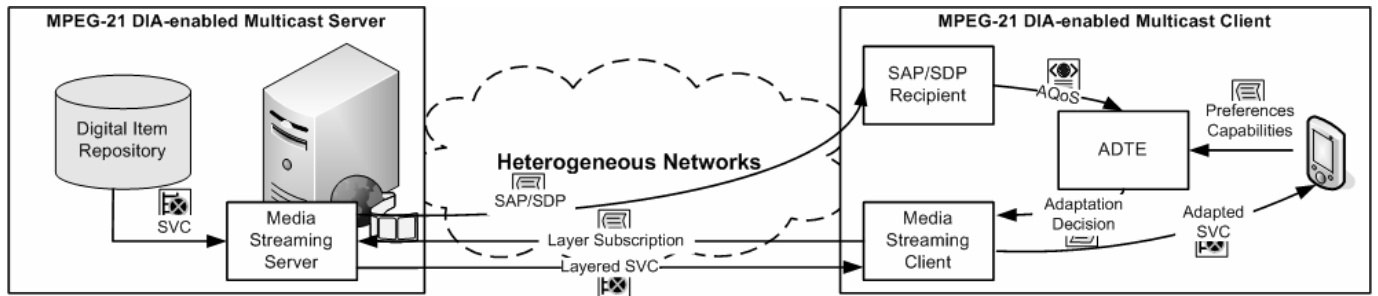


Figure 2: Interoperable Architecture for Multicast Streaming.

4 Evaluation

To evaluate the performance of our codec-independent metadata-based adaptation approach in comparison to an SVC-specific adaptation approach, the MPEG-21 DIA metadata-based implementation has been compared to the Bitstream Extractor of the JSVM reference software [4]. Initial comparisons have shown that the metadata-based adaptation approach clearly outperforms the implementation of the reference software [6]. However, these performance measurements have revealed bottlenecks for both, the metadata-based adaptation approach and the Bitstream Extractor. For the MPEG-21 DIA metadata-based adaptation approach, the transformation of the gBSDs utilizing an Extensible Stylesheet Language for Transformations (XSLT) library has been identified as the main bottleneck. For the Bitstream Extractor, the missing length information in the headers of the NAL units force the reference software to always parse the complete SVC bitstream, even if only a very small part of the bitstream needs to be extracted.

After identifying these bottlenecks, the metadata-based adaptation approach as well as the reference software were optimized [5]. For the metadata-based adaptation approach, a libxml-based transformation approach utilizing a generic transformation interface has been implemented. This transformation approach offers a clearly better performance than the XSLT libraries but still preserves the codec-independence of the MPEG-21 DIA metadata-based adaptation approach. For the Bitstream Extractor two optimizations have been implemented: Firstly, the Bitstream Extractor application of the reference software has been extended to utilize length information in the NAL unit headers. Those, only the NAL headers and not the complete bitstream need to be parsed. Secondly, an in terms of performance optimized Bitstream Extractor has been implemented, which illustrates the optimization potential for

the reference software as well as for the metadata-based adaptation approach.

In the following, the performance results for both, the optimized and the not optimized adaptation approaches are presented. The system for the performance measurements is a Dell Optiplex GX620 Desktop with an Intel Pentium D 2.8 GHz processor and 1024 MB RAM. As an operating system, OpenSUSE 10.2 was used and the performance measurements were made using OProfile System Profiler⁴ for Linux in version 0.93.

The performance measurements were performed for a number of bitstreams, due to space constraints only the results for two sequences, Mariposa and Ice, are presented. However, the measurements for all sequences have shown very similar results. The properties for both sequences are given in Table 1-2. The layer with the highest number represents the SVC bitstream in best quality, i.e. containing all lower layers, while the layers with a lower number provide a lower quality and are extracted by applying adaptation. The results of the performance measurements for Mariposa and Ice are shown in Figure 3 and Figure 4.

Firstly, the results confirm what has already been mentioned: When adaptation is desired, the MPEG-21 DIA metadata-based adaptation approach clearly outperforms the reference software. While the metadata-based adaptation approach works significantly faster if a lower layer is extracted and only a small part of the bitstream needs to be copied, the Bitstream Extractor always has to parse the complete bitstream and hence shows very similar results for the adaptation to all layers. When the results of the original implementations are compared with the optimized implementations, the performance has been improved by approximately one-third for both approaches. For the metadata-based adaptation approach, the improvement is due to the usage of the optimized transformation library, while for the reference software the optimization is caused by the usage of length information in the NAL headers.

Although the performance of both approaches has been optimized, the results of the customized Bitstream Extractor, which is an in terms of performance optimized implementation, show that there is still significant potential for further improvements. Depending on the sequence and the layer that should be extracted, the customized Bitstream Extractor is up to 80 times faster than the original Bitstream Extractor implementation of the reference software. Furthermore the customized Bitstream Extractor implementation shows that if the length information in the

	Layer 4	Layer 3	Layer 2	Layer 1	Layer 0
Resolution	4CIF	CIF	CIF	QCIF	QCIF
Framerate	15	30	15	30	15
Bitrate(kbps)	3264	1571	1228	376	281
Size (MB)	49	21.8	17	5.2	3.9

Table 1: Layer Properties of the Mariposa Sequence.

	Layer 3	Layer 2	Layer 1	Layer 0
Resolution	4CIF	4CIF	CIF	CIF
Framerate	30	15	30	15
Bitrate(kbps)	4512	3426	1005	773
Size (MB)	4.3	3.3	0.9	0.7

Table 2: Layer Properties of the Ice Sequence.

⁴ <http://oprofile.sourceforge.net/>

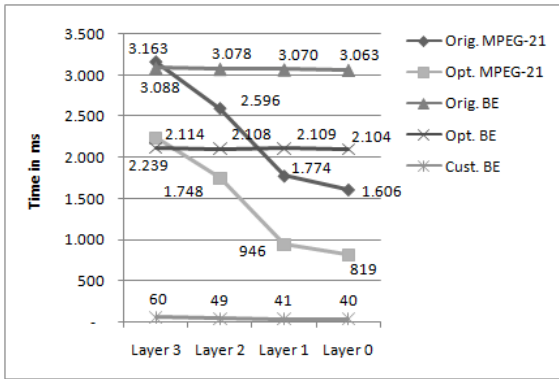


Figure 3: Mariposa Adaptation Performance

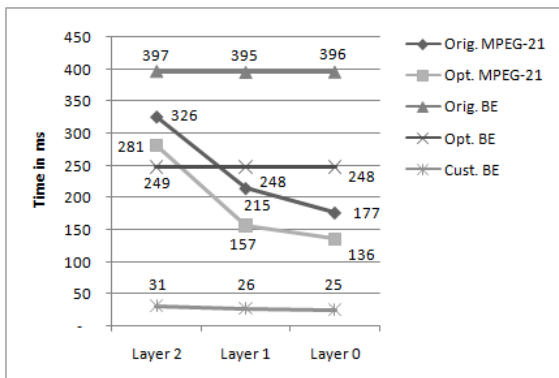


Figure 4: Ice Adaptation Performance

NAL headers is fully utilized, the extraction of lower layers can be performed clearly faster.

Although the customized Bitstream Extractor shows a very good performance, it should be noted that the application has been implemented for the sole purpose of adaptation and does not aim to provide the full functionality of the reference software's Bitstream Extractor.

In addition to the performance of the adaptation process, the overhead of the metadata in terms of storage size should also be taken into consideration. The gBSDs, which describe the structure of the bitstream on a high level and are used to perform the metadata-based adaptation process, have a size of 5-6 % of the bitstream's size for the tested sequences.

Overall, the evaluation results show that the MPEG-21 DIA metadata-based adaptation approach can very well compete with the implementation of the reference software. Furthermore, it has been shown that the optimizations for both approaches can significantly improve the performance of the adaptation process. However, the results of the customized and optimized Bitstream Extractor implementation show that there is still a lot of potential for further improvements.

5 Conclusion

In this paper, an interoperable streaming framework for VoD as well as multicasting is presented. The framework enables the streaming and adaptation of scalable multimedia content utilizing MPEG-21 DIA metadata for the adaptation purposes. On the one hand, the VoD architecture allows each client to set up a separate streaming session and to receive the

bitstream in a quality suiting its terminal's capabilities and the preferences of the user. To enable the server to adapt the bitstreams to the desired quality, the usage environment properties are provided to the server as MPEG-21 DIA metadata and the server takes the adaptation decision. The multicast scenario on the other hand reduces the network traffic by providing the same layered scalable multicast to all clients. In this scenario, the client takes the decision which layers to subscribe to based on the locally available usage environment properties and the information about the streaming session in the SAP/SDP announcements.

The performance evaluations have shown that the MPEG-21 DIA metadata-based adaptation approach outperforms the reference software's Bitstream Extractor when adaptation is desired. Furthermore, it has been shown that the metadata-based adaptation approach can be further optimized by using an optimized XML transformation library. The Bitstream Extractor has been optimized as well by using length information in the NAL headers. These optimizations have improved both adaptation approaches by approximately one-third. However, the performance of the customized Bitstream Extractor implementation shows that there is still a lot of potential for the optimization of both adaptation approaches.

References

- [1] B. Shen, W.-T. Tan, F. Huve, "Dynamic Video Transcoding in Mobile Environments", *IEEE MultiMedia*, vol. 15, no. 1, Jan.-Mar., 2008, pp. 42-51.
- [2] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", *IEEE Trans. on CSVT*, vol. 17, no. 9, September 2007, pp. 1103-1120.
- [3] A. Vetro, "MPEG-21 Digital Item Adaptation: Enabling Universal Multimedia Access", *IEEE MultiMedia*, vol. 11, no. 1, Jan.-Mar. 2004, pp. 84-87.
- [4] Joint Scalable Video Model (JSVM) 9, Joint Video Team (JVT) of ISO/IEC and MPEG & ITU-T VCEG, N8751, Marrakech, Morocco, January 2007.
- [5] M. Eberhard, L. Celetto, C. Timmerer, E. Quacchio, and H. Hellwagner, F. Rovati, "An Interoperable Streaming Framework for Scalable Video Coding based on MPEG-21", *Proc. VIE 2008*, Xi'an, China, July 2008.
- [6] M. Eberhard, L. Celetto, C. Timmerer, E. Quacchio, and H. Hellwagner, "Performance Analysis of Scalable Video Adaptation: Generic versus Specific Approach", *Proc. WIAMIS 2008*, Klagenfurt, Austria, May 2008.