REAL-TIME DVB-BASED MPEG-21 DIGITAL ITEM ADAPTATION FOR LIVE UNIVERSAL MULTIMEDIA ACCESS

Martin Prangl, Christian Timmerer and Hermann Hellwagner

Department of Information Technology (ITEC), Klagenfurt University
{firstname.lastname}@itec.uni-klu.ac.at

Department of Information Technology (ITEC)
Klagenfurt University
Technical Report No. TR/ITEC/06/1.04
June 2006

Real-time DVB-based MPEG-21 Digital Item Adaptation for live Universal Multimedia Access

Martin Prangl, Christian Timmerer and Hermann Hellwagner
Department of Information Technology (ITEC), Klagenfurt University, Klagenfurt, Austria
E-mail: *{firstname.lastname}@itec.uni-klu.ac.at*

Abstract - *In order to enable transparent and augmented use of multimedia content across a wide range of networks and devices, content adaptation is an important issue within multimedia frameworks. In this paper, we present a prototype application that receives Digital Video Broadcast (DVB) TV streams on a PC, transcodes the streams on the fly according to the individual User requirements and packs the adapted content together with available metadata into a standard compliant MPEG-21 Digital Item (DI). In this form, the framework enables the live Universal Multimedia Access (UMA) scenario where the DVB content can be transparently accessed by clients such as PCs and PDAs, anytime and anywhere.*

Keywords - *DVB, MEPG-2, MPEG-4, MPEG-7, MPEG-21, Digital Items, metadata*

1. INTRODUCTION

Television (TV) content is usually broadcasted by air and consumed with special purpose devices such as TV sets which are connected to special antennas in order to receive this broadcast service. However, TV content delivered over best effort networks, like the internet are becoming increasingly popular. So called Internet Protocol Television (IPTV) services delivering high quality TV channels to the user's living room are becoming more and more attractive because they are able to receive these digital channels on their regular PCs or any other Internet-enabled device. However, due to the Internet's best effort characteristics continuous bandwidth, delay, or jitter cannot be guaranteed between heterogeneous network nodes. Such guarantees are only possible – if at all – on certain parts of the link between the provider and user's end device. This issue results in two major drawbacks. First, the use of lightweight terminals such as personal digital assistants (PDAs) or handhelds are limited due to the fluctuating network characteristics. Second, the capabilities of these devices, i.e., memory, computational power, or display resolution may be diverse as well.

Therefore, one major goal within the multimedia research community is the development of Universal Multimedia Access (UMA) [1] strategies and technologies which enable users to consume any kind of multimedia content, anywhere, and anytime. To achieve this goal, the multimedia content has to be adapted to meet the limitations of the user's terminal and network characteristics. Such multimedia adaptation could be, e.g., transcoding from one video format to another or scaling a video in the spatial domain such that it fits on the terminal's screen. Furthermore, the content itself must also be adapted such that a user has an informative experience; in other words, the "end point" of universal multimedia consumption is the end user and not the terminal. Therefore, enabling the vision of Universal Multimedia Experience (UME) [2] might include, e.g., insertion of subtitles into a video allowing deaf users to follow the spoken content in a video.

In this paper we present a framework that is able to receive DVB[1] streams from satellite and transcode the content in real-time individual to the client's terminal capabilities and the user's demands. The proposed framework delivers the adapted content over an IP-based network to which the end user is connected. This enables users to consume their preferred DVB service independent of the actual satellite coverage. In order to enable this kind of UMA, both the transcoded multimedia content and its associated metadata are packed into a MPEG-21 Digital Item (DI) providing an open and interoperable interface to the multimedia content. In contrast to our prior work [3] where we focused on the metadata extraction itself, in this paper we focus on real-time transcoding of the DVB content into various output formats (e.g., MPEG-1, MPEG-4, etc.), the transformation of the corresponding metadata, and the final delivery to the end user.

The remainder of this paper is organized as follows. Section 2 provides an overview of the whole system and briefly discusses the main functionalities of each module. The content aware adaptation step is described in Section 3. Thus, the associated metadata which needs to reflect the modified content has to be adapted as well. This step is outlined in Section 4. Section 5 deals with the delivery of the adapted content to the client and Section 6 concludes the paper.

[1] http://www.dvb.org

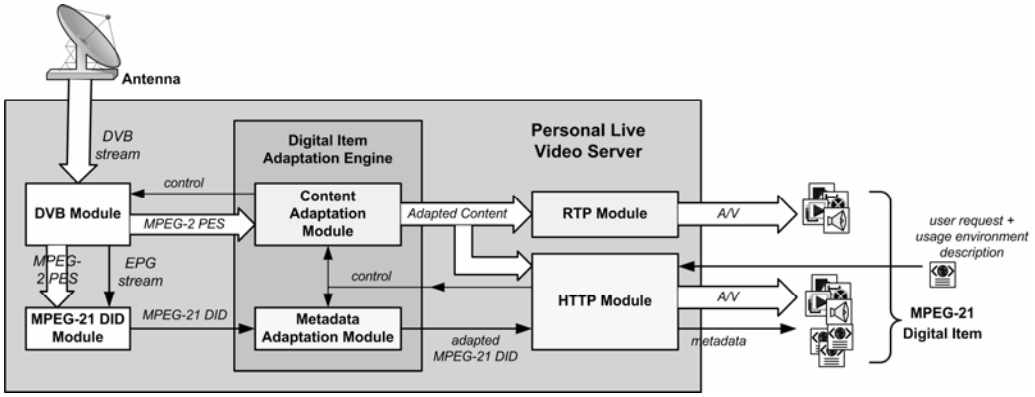

Figure 1 — *dvb2di* (DVB to Digital Item) system architecture.

2. SYSTEM OVERVIEW

The overview of the proposed adaptive *dvb2di* (DVB to Digital Item) framework is depicted in Figure 1. The *DVB module* receives the DVB input from the DVB hardware. It is responsible for tuning to the right transponder frequency, de-multiplexing the MPEG-2 transport stream and extracting the electronic program guide (EPG). The DVB module is controlled by the adaptation engine which receives only selected audio-visual (A/V) streams for further processing. The EPG and MPEG-2 Packetized Elementary Stream (PES) is forwarded to the *MPEG-21 DID module*. This module takes the EPG and the low-level stream information which is used to form a MPEG-21 Digital Item Declaration (DID). The DID is an XML-based description which provides means for associating (multimedia) content and metadata within an interoperable structure for delivery and exchange among users.

The central module of the system is the so called *DI adaptation engine* which can be divided into two parts. First, the *content adaptation module* takes the MPEG-2 PES stream containing the audio and video information of the selected channel and adapts or transcodes it to the user's terminal capabilities and/or preferences (e.g., changing the spatial resolution and/or the coding format). Second, the *metadata adaptation module* takes the DID of the original content and transforms it accurately according to the content features of the adapted A/V stream. This step is necessary because, if it is omitted, the wrong syntactic (and semantic) description of the adapted media stream would be delivered.

The *HTTP module* takes as input the user's request for a certain multimedia content including the user's preferences and terminal capabilities. A Web interface presents the available channels and a short description thereof to the user. According to the usage environment description the HTTP module checks if the framework can perform the request and if so, it configures the adaptation engine correspondingly. In the current implementation the user is appropriately informed in case the request cannot be fulfilled. As an answer to the successful request, the HTTP module provides the adapted the DID for consumption. In this DID the URL of the adapted live A/V stream is given as well as corresponding metadata. The system implements two options – based on the user's choice – for delivering the requested content over the IP network, i.e., the Real-time Transport Protocol (RTP) and the Hypertext Transfer Protocol (HTTP). Both means for transportation are further detailed in Section 5.

3. CONTENT ADAPTATION

For content adaptation it is important having knowledge of the context where the content is ultimately consumed. In our framework, this context information, namely all hardware and software capabilities including the resource limitations of such end devices as well as the individual user preferences (e.g., impairments) are described with MPEG-21 DIA Usage Environment Descriptions (UEDs) [4]. Having this UED of the requesting client, the provider has to choose the right adaptation decisions in order to deliver an adapted variation of the requested content which fits the terminal's capabilities and its resource limitations. These decisions are provided by an adaptation decision-taking module which is, in our case, a simple table look-up. Enhanced decision strategies are preserved for future work and not further detailed here due to space restrictions.

The resulting adaptation parameters are forwarded to the adaptation engine which performs the actual multimedia content adaptation. The current content adaptation module supports currently three scalability dimensions (i.e., spatial, temporal, and signal-to-noise ration (SNR)) and takes the DVB content in form of an MPEG-2 PES comprising the elementary A/V streams. First, this container format has to be de-multiplexed into the A/V elementary streams (ES). Each extracted ES may be decoded into the *uncompressed domain* for further manipulation. The video ES is decoded in the YUV format allowing for spatial and SNR adaptability. The former is used to reduce the resolution of each frame, e.g., from 640x480 to 320x200 pixels

whereas the latter enables the reduction of the quality for each frame by applying individual quantization parameters during the encoding step. in case no spatial or SNR adaptability is needed, i.e., the target codec is the same as the source codec, no decoding of the video ES is required, which leads to a much lower CPU consumption. A possible adaptation in this *compressed domain* is called temporal adaptation, i.e., the dropping of individual frames from the original video stream.

If the audio ES has to be adapted, it will be decoded into the pulse-code modulation (PCM) format. In the decompressed domain the audio signals can be re-, down-sampled, or the number of available channels can be reduced (e.g., stereo to mono).

Following target A/V codecs are supported by our framework:

- MPEG-1 Visual
- MPEG-1 Audio @ Layer 2 (mp2)
- MPEG-1 Audio @ Layer 3 (mp3)
- MPEG-2 Visual
- MPEG-4 Advanced Video Coding (AVC)
- MPEG-4 Advanced Audio Coding (AAC)

Table 1 shows the average CPU consumption of the trancoding process for transcoding a single high-motion action video (approx. 60 sec.) into several degraded variations. Column *q* represents the applied DCT quantization parameter of each variation. The target codecs were set to MPEG-2 Visual and MPEG-4 AVC respectively. Additionally, the corresponding average bandwidth (BW) of the produced content variations is shown. The original bitstream has a spatial resolution of 720x576 pixels and a frame rate of 25 fps with an average bit-rate of 5,800 kBit/s. It can be noticed that the CPU consumption which is needed for transcoding to MPEG-4 AVC is much higher than adapting the MPEG-2 stream. The highest reachable frame rate was determined by 16 fps for real time transcoding into MPEG-4 AVC, maintaining the original spatial resolution and does not exceed the CPU power limitations. Taking the other option, i.e., maintaining the original frame rate, the maximum resolution is 500x400 pixels. The hardware for these experiments was a Pentium D processor with 3.0 GHz and 1 GB RAM. The implementation of the content adaptation module includes *ffmpeg*[2] library calls enabling efficient de-multiplexing, decoding and encoding of various A/V codec types.

Table 1 — Performance of video ES transcoding.

Codec type	Spat. Res	Fr [fps]	q	CPU [%]	BW [kBit/s]
MPEG-2	720x576	25	5	35	2,600
MPEG-2	720x576	25	10	15	1,750
MPEG-2	360x288	25	1	20	1,400
MPEG-2	360x288	25	5	16	750
MPEG-2	360x288	25	10	14	560
MPEG-4	720x576	16	1	99	820
MPEG-4	500x400	25	1	99	760
MPEG-4	360x288	25	1	25	380

4. METADATA ADAPTATION

In order to reflect the changes applied to the multimedia content, the corresponding metadata needs to be adapted as well. In our dvb2di framework this task is done by the metadata adaptation module. This modules takes the DID from the MPEG-21 DID module as an input and provides an adapted DID as output. It is controlled by the ADTE located within the HTTP module and actually the same parameters as for the content adaptation engine are used for adapting the metadata. Document 2 shows the media information of the original Digital Item (i.e., before the adaptation step) as an MPEG-7 description which is part of the DID (cf. Document 1) and associated to the multimedia content.

The MPEG-7 media information provides means for describing syntactical aspects of the multimedia content such as the coding format, frame size/rate or sample rate. Our media information description can be categorized into three parts. First, general information about the type of the content and its bit-rate are provided within the `Content` element. Second, the `VisualCoding` element describes the coding format as well as frame size and frame rate of the visual part of the A/V content. In our case, the source video is encoded as an MPEG-2 Video with a frame size of 720x576 pixels at 25 fps. Finally, the audio part is described within the `AudioCoding` element in terms of its coding format, audio channels, sample rate and the audio sample accuracy in bits per sample.

Due to the small size of the current MPEG-21 DID, the actual metadata adaptation is performed by an XSLT[3] module within the metadata adaptation module. The parameters are the same as for the content adaptation module, i.e., the new frame size/rate, coding format, number of audio channels, etc. In case the size of overall DID is increasing, possibly due to further semantic or syntactic annotations extracted from the EPG, it is envisaged to incorporate more memory efficient XML transformation techniques such as Streaming Transformations for XML (STX) [5].

5. CONTENT DELIVERY

In order to deliver the live A/V content over best effort networks, the specific network protocol characteristics of available protocols have to be considered, enabling continuous consumption of the (live) content. Due to the fact that the Internet is an

[2] http://ffmpeg.sourceforge.net/

[3] Extensible Stylesheet Language for Transformations;

```
<MediaInformation id="S19.2E-0-12480-899">
 <MediaProfile><MediaFormat>
  <Content href="MPEG7ContentCS:2001">
   <Name>audiovisual</Name>
  </Content>
  <BitRate average="3500000"
    maximum="5000000">3000000</BitRate>
  <VisualCoding>
   <Format href="urn:mpeg:mpeg7:cs:
      VisualCodingFormatCS:2001:2"
      colorDomain="color">
    <Name xml:lang="en">MPEG-2 Video</Name>
   </Format>
   <Frame height="576" width="720"
      rate="25.0"/>
  </VisualCoding>
  <AudioCoding>
   <Format href="urn:mpeg:mpeg7:cs:
      AudioCodingFormatCS:2001:3.2">
    <Name xml:lang="en">
      MPEG-1 Audio Layer II</Name>
   </Format>
   <AudioChannels front="0" side="2"
      rear="0" lfe="0" track="0">2
   </AudioChannels>
   <Sample rate="48000.0" bitsPer="192"/>
  </AudioCoding>
 </MediaFormat></MediaProfile>
</MediaInformation>
```

Document 2— MPEG-7 media information of orginal DI

IP-based network we have two choices for delivering the A/V content. The first choice is to use RTP/UDP for delivering each ES to the client. Each ES is paketised into RTP packets according to the codec specific RFCs, e.g., each frame in one packet. This packetization scheme allows for a certain packet loss which results from decreasing bandwidth, i.e., the client would receive not all frames but is still able consuming the content without having to wait for retransmitted packets.

However, in some scenarios the use of RTP is not desired due to network management issues – such as firewalls which may block UDP traffic – which increase the demand for alternative transmission techniques. A possible answer to these issues is the use of HTTP for continuous media transmission. The choice between RTP- and HTTP-based transmission is implemented in our framework and signalled to the client through the DID's choice/selection mechanism as shown in Document 1. This additional option for content streaming, also kown as progressive download, is provided by the HTTP module. It receives a HTTP-GET request for the selected content and transmitts the adapted A/V bitstream, which is packed in the PES format, as answer to the client. However, this delivery method is limited to unicast scenarios which is a widespread use case for UMA-enabled applications. Major drawbacks of multimedia content delivery over HTTP form the TCP-specific features like conquestion and flow control. In case of dynamic bandwith limitations during the delivery, the live consumption may stuck caused by these mechanisms. Nonetheless, because TCP is a lossless protocol, we know exactly – at the server-side – how

```
<DIDL><Item>
 <Choice> <!-- choice/selection for user -->
   <Selection select_id="http"/>
   <Selection select_id="rtp"/>
 </Choice>
 <Component>
  <Condition require="http"/>
  <Descriptor><Statement>
    <!-- media information (Document 2) -->
   </Statement></Descriptor>
  <Resource ref="http://www.a.com/media"/>
 </Component>
 <Component>
  <Condition require="rtp"/>
  <Descriptor><Statement>
    <!-- media information (Document 2) -->
   </Statement></Descriptor>
  <Resource ref="rtp://www.a.com/media"/>
 </Component>
</Item></DIDL>
```

Document 1— MPEG-21 Digital Item Declaration

many bytes were deliverd to the client within a given time period. This knowledge let us adjust the adaptation parameters on the fly. In case of decreasing bandwidth we are able to react with a lower frame rate or higher quantisation resulting in lower a bit rate of the live A/V stream. However, the quality of the A/V stream decreases (as in case of RTP) but the consumption keeps continuously.

6. CONCLUSION

In this paper we have presented the dvbtodi framework which enables transparent and MPEG-21 conformant access to DVB content taking into account the usage environment constraints where the content is ultimately consumed. Currently we are improving the adaptation decision-taking component towards sofisticated decision strategies including modality conversions.

7. REFERENCES

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

[2] F. Pereira and I. Burnett, “Universal Multimedia Experiences for Tomorrow,” IEEE Signal Processing Magazine, vol. 20, no. 2, 2003, pp. 63-73.

[3] M. Prangl, C. Timmerer, K. Leopold and H. Hellwagner, “DVB-based MPEG-21 Digital Items for Adaptive Multimedia Streaming”, *Proc. of the 47th International Symposium ELMAR-2005 focused on Multimedia Systems and Applications*, Zadar, Croatia, June 2005.

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

[5] O. Becker, "Transforming XML on the fly", in XML Europe 2003, May 2003.