

An MPEG-21-enabled Video Adaptation Engine for Universal IPTV Access

Ismail Djama

CNRS LaBRI Lab. –
University of Bordeaux 1,
Talence, France
djama@labri.fr

Toufik Ahmed

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

Daniel Negru

CNRS PRiSM Lab –
University of Versailles,
Versailles, France
dan@prism.uvsq.fr

Nawel Zangar

CNRS PRiSM Lab –
University of Versailles,
Versailles, France
zana@prism.uvsq.fr

Abstract

In the context of Universal Multimedia Access (UMA), the user will be able to access and consume a rich set of multimedia content over dynamic and heterogeneous networks and devices. To reach this universal accessibility, the MPEG-21 multimedia framework is defining the entire multimedia content delivery chain (content creation, production, delivery, trade and consumption). The part 7 of this framework defines the digital item adaptation (DIA). The DIA specifies metadata for assisting the adaptation of digital item according to different constraint (storage, transmission, consumption). This paper presents an architecture and implementation of an MPEG-21-enabled Video Adaptation Engine (VAE) for allowing seamless IPTV access for heterogeneous devices and networks capabilities. The IPTV services considered in this paper are based on streaming Digital Video Broadcast –Terrestrial services over IP LAN and WLAN networks. The VAE is seen as a media gateway proxy with adaptation/streaming capabilities. The obtained results demonstrate the ability of our VAE to allow universal media access to digital content for a multitude of clients and network capabilities.

Keyword

IPTV, DVB-T, Adaptation engine, MPEG-21 Digital Item Adaptation.

INTRODUCTION

Within the last few years, the tendency has been towards convergence between the three worlds of network communications (data, telephony, and broadcast). This network convergence is aimed to enable the Universal Media Access (UMA) through which the user will be able to access anytime and anywhere any services using different access networks and devices.

Recently the All-IP concept has emerged as the promising concept to allow this convergence. In fact, in the All-IP approach, a common IP network layer is introduced in each different communication network aiming to realise the interconnection. This All-IP approach is reflected in the recent developments on broadband access. Actually, many operators propose multi-play services that allow customers to simultaneously access Internet, telephony based on Voice over IP (VoIP), and television based on IPTV. However, the All-IP approach must deal with

different open issues inherited from the IP protocol: heterogeneity, end-to-end QoS, mobility, and security.

In the context of heterogeneity, the services proposed to the users (web, mail, telephony, television...) must be customized according to the networks, user and terminals configuration.

In this article, we focus on IPTV services where Digital TV is transmitted over different IP networks. In particular, we investigate the delivery of DVB (Digital Video Broadcast) programs over IP. The DVB project [1] is an industry-led consortium that regroups broadcasters, manufacturers, network operators, software developers, regulatory bodies and others in over 35 countries committed to designing global standards for the global delivery of digital television and data services. The DVB project defines different standards for transmission of digital television: the DVB-S (satellite transmission), DVB-C (cable transmission), and DVB-T (terrestrial transmission). Recently, the DVB project has defined the DVB-H (handheld) standards which adapt DVB-T to the specific requirements of handheld, battery-powered receivers, and user mobility. However, the DVB-H remains a broadcast transmission and misses an important component for services interactivity. The introduction of the internet protocol (IP) and an uplink transmission (wireless, 3G) is imperative for services interactivity.

Allowing TV reception through heterogeneous IP environment is very challenging. The TV stream must be adapted to different constraints (user preferences, network configuration, and terminal capabilities, etc.). In general, audiovisual applications are bandwidth adaptive but have stringent delay, jitter, and packet loss requirements. Consequently, one of the major requirements for the successful and wide deployment of such services is the efficient transmission of sensitive content (audio, video, and image) over a broad range of bandwidth-constrained IP access networks.

For this purpose, we propose an MPEG-21-enabled Video Adaptation Engine (VAE) for allowing seamless IPTV access for heterogeneous devices and networks capabilities. Our VAE is based on MPEG-21 multimedia framework part-7 which defines the Digital Item Adaptation (DIA) concept.

MPEG-21 [2] aims to define a normative open framework to support users to exchange, access, consume, trade and otherwise manipulate digital item (DI) in an efficient, transparent and interoperable way.

The framework is organised in different parts. The part-7 [3] defines the Digital Item Adaptation (DIA) concept which specifies normative descriptions of tools to assist with the adaptation of digital Items. We extend the VAE with SDPng (Session Description Protocol Next Generation) protocol [4] to carry the metadata information related to the environment and the user constraints.

The remainder of this paper is organised as follow: section 2 gives an overview of the MPEG-21 framework and Universal Media Access (UMA). Section 3 presents the proposed IPTV platform architecture. In section 4, we present some implementation issues followed by performance evaluation of the platform in section 5. We conclude and give some perspectives for future work in section 6.

MPEG-21 AND UMA CONCEPT

MPEG-21 [2] is the newest standards series being produced by the Moving Pictures Experts Group. The MPEG-21 vision aims to define multimedia framework to enable transparent and interoperable use of multimedia resources across heterogeneous networks and devices. MPEG-21 is based on two principal concepts: the digital item (DI) and the user's interactions with this digital item. The DIs recover all the media content type (video, audio, animation, image...). The interactions represent all the actions that the users can perform on media content (generation, trading, exchanging, consuming...).

The MPEG-21 standards are organized into seven independent parts in order to facilitate their usage and implementation.

On the other hand, the UMA (Universal Media Access) system aims at enabling access to multimedia content anytime, anywhere and without any concern on particular formats, devices, networks. The UMA deals with the delivery of media resources under different network conditions, user preferences, and capabilities of terminal devices.

It is obvious that there is a match between the goals of UMA and MPEG-21. However, it may be necessary to adapt the DIs according to the actual usage environment and context configuration.

This adaptation is defined in MPEG-21 part 7: Digital item adaptation (DIA) [3]. The DIA specifies tools to assist with the adaptation of Digital Items (DIs). These different tools are based on metadata using the eXtensible Markup Language (XML). The Usage Environment Description (UED) constitutes the main tool. It provides a means for describing

terminal capabilities, network capabilities, delivery capabilities, and user characteristics...

Terminal capabilities include encoding, decoding capabilities, display and audio output capabilities, and other hardware characteristics (processor speed, power consumption, memory capacities). Network capabilities include static capabilities such as maximum throughput as well as dynamic conditions such as bandwidth variation, loss ratio, and delay. Delivery capabilities recover the signalisation and transport protocols supported. User characteristics represent the user preferences such as language, media format, and dissemination mode.

Three others tools are used by DIA to optimize the adaptation: gBSD (generic Bit Stream Description), UCD (User Constraint Description), and AQoS (Adaptive Quality of service). The gBSD represent a high-level description of DI (for examples: frames, packets). The UCD specifies constraints on the user environment. Finally, the AQoS gives more precision for how to perform the adaptation according to constraints and keeping some quality of service.

Figure 1 illustrates the general framework of MPEG-21 DIA. However, the adaptation engine is left open to various implementations.

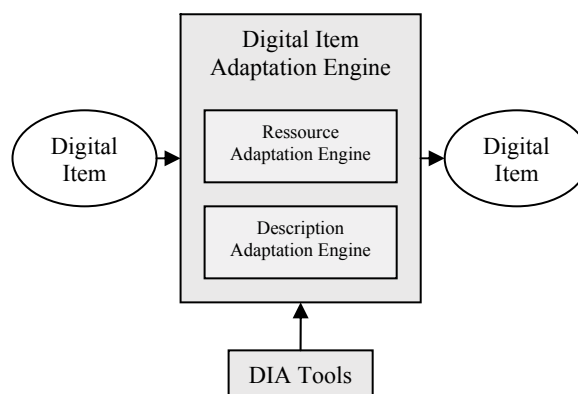


Figure 1: MPEG-21 DIA Framework

NETWORK CONVERGENCE: IPTV CONCEPT

The convergence of data, telephony, and broadcast networks remained a dream during several years for many providers. New generations of wired and wireless networks (ADSL, cable, UMTS, WIFI, DVB-H/T...) have been recently widely accepted for accessing different IP-based service. These new networks offer access to different services (telephony, internet) and different content media (VoD, television). IP protocol is playing an important role to allow service convergence. It aims to assure end-to-end interconnection and thus building of an efficient and simple network that regroups heterogeneous networks and services.

However, to reach this ultimate goal, some issues in the IP protocol must be addressed such as QoS, mobility, and security.

The QoS represents the first issue for the IP protocol for carrying the multimedia streams. In fact, assuring end-to-end QoS for audiovisual services across different heterogeneous networks is very challenging. The media stream will face the bandwidth variation, packet losses, jitter, and delay problems.

At the present time, there are some solutions to enhance the QoS in IP networks. To manage the bandwidth, different techniques coexist such as stream switching [5][6][7], simulcast, transcoding [8][9][10], and congestion control algorithms [11][12]. To deal with packet losses, mechanisms such as end-to-end retransmission, Forwarding Error Correction (FEC) [13][14], and Application level Framing for error resilience have been proposed. The jitter is absorbed by using a client buffer of which the size must be optimally determined. The transmission delay is very important for interactive services (VoIP, videoconference). To minimize this delay, different QoS-enabled network mechanisms are deployed in the core of the network such as IntServ (Integrated Services) with its resource reservation protocol RSVP [15] and Diffserv (Differentiated services) to provide different packets treatments [16].

IPTV consists of new types of services proposed in next generation networks. The IPTV system consists of digital television to be delivered to subscribing users using the IP protocol over different possible connections (ADSL, cable, WiFi, DVB-T/H). It permits to cover live TV (broadcasting) and VoD (Video on Demand) services. Different services can be developed using IPTV comparing with the tradition TV broadcasting. Among them, we can find: pay-per-view services, interactive TV and scheduled program retransmission. Moreover, the TV stream can be customized according to the networks and terminals access and user preferences (screen format, movie's languages...). IPTV will also permit the mobility of TV. The user will be able to watch TV anytime and anywhere and

access his TV subscriptions even if he is not at home.

PLATFORM ARCHITECTURE

It is important to design IPTV service capable to adapt to system and network resource constraints while ensuring that end-user requirements are met. Figure 2 shows a general framework of IPTV adaptation. It is composed of three components: the content server, the video adaptation engine (VAE), and the client. In our architecture, all the adaptations are performed in the VAE which is represented as an independent node in the network. In fact, performing content adaptation in an independent node seems to be the prominent solution compared to performing the adaptation at the client or at the content server side. The client resources can be limited and the content server can be overloaded by serving several clients at the same time. The content server is responsible of receiving the DVB-T stream which is encoded as an MPEG-2 Transport stream (TS). The client provides metadata on the session and on individual components and requests the TV stream using different signalisation protocols (HTTP/RTSP/MMS).

To carry the usage environment description defined by MPEG-21, the client uses SDPng (Session Description Protocol - Next Generation) [4]. The SDPng is the successor of SDP. It is being currently defined within the MMUSIC (Multipart Multimedia Session Control) working group of the IETF.

While the SDP consists of a number of lines of text of the form <type>=<value>, the SDPng is based on XML that allows more flexibility and extensibility. To express more complex and then richer media features and scenarios, the SDPng is organised in five description sections: (1) capabilities used in the negotiation process to match the capabilities of the current users, (2) definitions provides commonly used parameters, (3) configurations describes the different session components, (4) constraints section gives some constraints on relation with configurations, and (5) session information provides metadata on the session and on individual components.

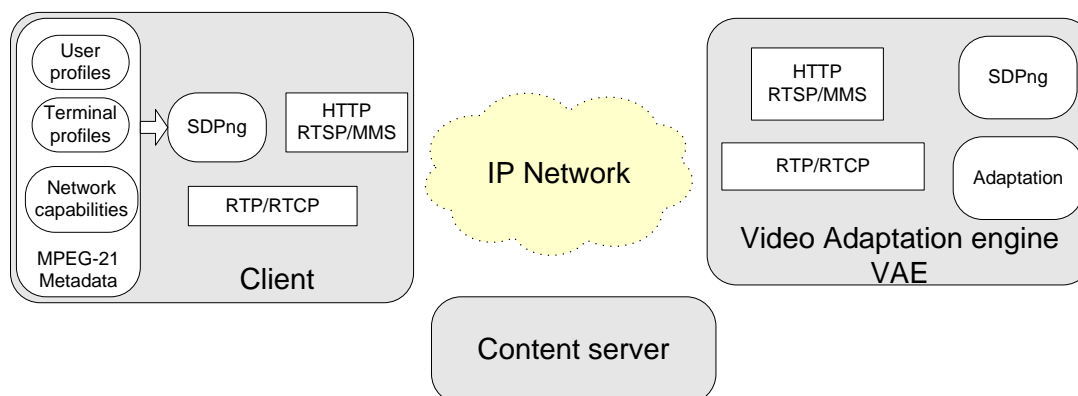


Figure 2: General framework of IPTV adaptation

Recently, harmonization between MPEG-21 and SDPng has been proposed in [16][17] and [18] in the packages standardization process.

Based on this harmonization, we elaborate an SDPng document that respects the global SDPng structure and includes in the different sections simplified descriptions from the MPEG-21 UED.

The SDPng Document is carried using the signalisation protocols (RTSP/MMS/HTTP).

When the VAE receives a request from the client, it extracts the SDPng document and parses it. From this parsing, the VAE is initialised with different adaptation features (codecs, video bitrate, screen size...). In the same time, it requests the TV stream from the content server. Thus, the TV stream received by the adaptation engine is adapted and sent out to the client using RTP (Real-time Transport Protocol).

PLATFORM IMPLEMENTATION

The proposed platform components are implemented based on open source project VideoLAN [19]. The VideoLAN project is a complete software solution for video streaming and playback. It is based mainly on VLC software which offers different functionalities: media player, media server, and real time adaptation functionalities. Moreover, it supports different codec formats (MPEG-1, MPEG-2, MPEG-4, DivX, mp3, ogg...) and different streaming protocols (UDP Unicast, UDP Multicast, RTP, HTTP, RTSP, and MMS).

In the content server, VLC is used to capture the DVB-T stream using an USB DVB-T card. We have used FRECOM DVB-T USB Stick [20]. The client and the VAE use a modified version of VLC which is augmented with SDPng functionalities, and adaptation capabilities to meet the user's requirements.

The SDPng draft proposes different usage of SDPng in different applications scenarios. In our implementation, we were interested on carrying the SDPng document over RTSP (Real Time Streaming Protocol).

However, to adapt this transport protocol to our needs, we follow a two-way session setup. In fact, in the RTSP suggestion the server provides a complete SDPng document describing all components making up a presentation and includes detailed codec and transport parameters for each of them. The client may only pick one out of alternatives for each of the offered components but has no further option to negotiate parameters in depth. This is not conformed to our needs since the client has to send a description of his capabilities to the VAE.

It is worth noting that RTSP does not propose a complete Client/Server capabilities negotiation. The original SDP document is send by the server to the client. Thus, we used the two-way session setup

such as proposed by SIP (Session Initiation Protocol) and we adapted it to RTSP [21] negotiation. Figure 3 summarises the interaction between the client and the VAE using RTSP with a two-way session setup. This setup is detailed in the following:

After exchanging different supported commands by the two parts with the OPTION command, the client sends a DESCRIBE command that includes in its body the required client SDPng document to the VAE. The VAE parses the client SDPng document and generates its own SDPng document according to the client capabilities. The SDPng document of the VAE is sent to the client in the DESCRIBE response body. Thus, both sides now know the common configurations and can initiate the session.

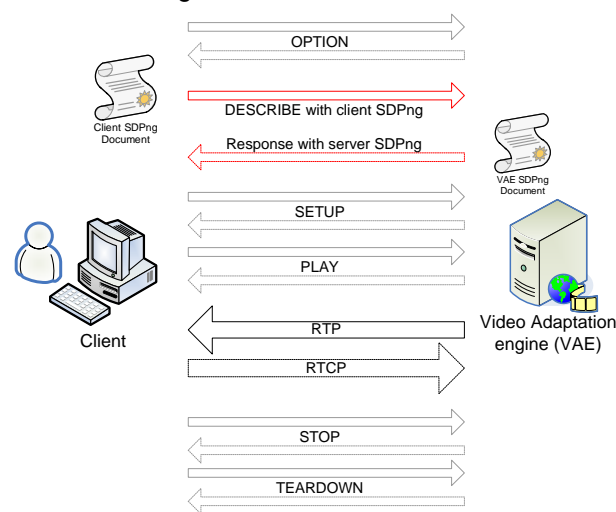


Figure 3: Client/adaptation engine interaction

The content server receives the MPEG-2 TS stream from the USB DVB-T Stick. The TS packets are encapsulated without any transformation in UDP packets which are directly transmitted to the adaptation engine. In case of Ethernet based networks, one UDP packet carries seven TS packets.

The VAE tries to adapt the TV stream obtained from the content server according to the client SDPng document which describes its capabilities. The adaptation performed by our adaptation engine is a structural level adaptation. In fact, the adaptation engine can transcode (change the video or audio codec), transrate (change the video or audio bitrate), resize the video frame, and change the frame rate of the TV stream and all this in real time. This adaptation is based on ffmpeg [22] API of VLC.

PLATFORM OPERATION EXPERIMENTATIONS

To demonstrate the operation of our MPEG-21 enabled Video Adaptation Engine, we have conducted different experimentations.

Figure 4 shows our testbed topology. We used wired (Ethernet) and wireless (802.11b) networks and three types of access terminals: an office computer with wired connection, a laptop and PDA (Personal Digital Assistant) both of them with wireless connections.

To limit the available bandwidth in the two networks, we used a Linux router with TC command line interface provided by Linux iproute2 package.

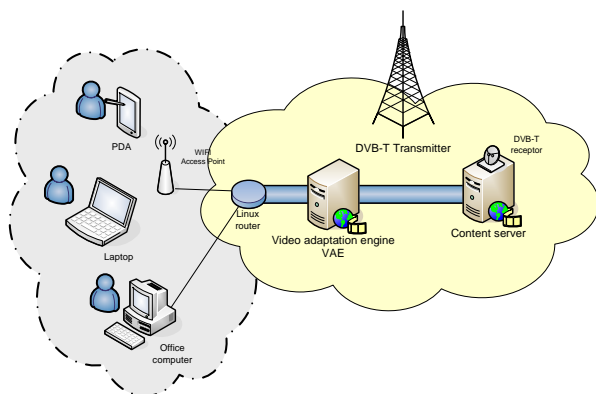


Figure 4: IPTV Testbed topology

The MPEG-2 video stream captured in the content server has 720×576 pixels size, an average throughput of 1913 Kbps, a peak rate of 2565 Kbps, and frame rate of 25 frames / second. The audio (MPEG-II Audio) throughput is 192 Kbps.

We fixed different characteristics supported for the three clients:

Office computer client: with MPEG-1 codec for video and MPEG1 Layer 1 audio codec. Network throughput is limited to 2048. The screen size is 720×576 pixels.

Laptop client: with WMV / WMA v1 codec for video and audio. Network throughput is limited to 1024 Kbps. The screen size is 512×384 pixels. In this scenario, we supposed that only TCP protocol is being supported by the client.

PDA client: with MPEG-4 codec for video and 15 frames/seconds supported. Audio uses MPEG Layer 3 codec. Network throughput is limited to 512 Kbps. The screen size is 320×240 pixels.

For each client, The VAE receives the corresponding profiles / characteristics and determines the best video and audio bitrates according to the network throughput and the screen

size. The VAE then transcodes and transmits in real time the audio and video stream according to the codec and transport protocol supported by the client.

The reduction of frame rate (e.g. in case of PDA client) is performed by dropping some frames according to hierarchies B-Frame, P-Frame and I-Frame. In fact, the I-frame is more important than P-frame which is more important than B-frame.

(a) (b) (c)

Figure 5 shows some video snapshot examples. The Figure (a) shows the original video frame, Figure (b) the video frame perceived at the laptop client without VAE adaptation and Figure (c) the video frame perceived at the laptop client with VAE adaptation. Although there is a quality drop between Figure (a) and Figure (c), the laptop client prefers to perceive the Figure (c) video quality compared to Figure (b) in which some packet losses drastically affect the perceived quality.

CONCLUSION

To reach the ultimate goal of UMA (Universal Media Access), we presented in this paper an architecture and an implementation of an MPEG-21-enabled Video Adaptation Engine (VAE) for allowing seamless IPTV access to heterogeneous devices with respect to networks capabilities. IPTV certainly represents the future of TV transmission since it enables the deployment of new services (VOD, interactive program) that were been impossible in the past, due to the lack of interactivity.

Our system is focused on enabling the anytime, anywhere, anyhow TV access through the integration of user/terminal characteristics, content descriptions, and networks capabilities.

Our architecture has been designed according to the recent multimedia standards: MPEG-21 part 7 and SDPng protocol.

The different experimentations conducted validate our architecture and prove the requirement of the adaptation at the connection phase.

Future works are focused on long-term adaptation. This consists of the adaptation during the stream transmission phase allowing the server to react to the dynamic changes in characteristics: bandwidth in the case of network congestion and terminal access in the case of mobility. Paper [23] addresses this problem and presents a cross-layer system for adapting audiovisual content according to network congestion and end-to-end requirements.



(a)



(b)



(c)

Figure 5 Quality comparison for perceived video at laptop client

REFERENCE

- [1] Digital Video Broadcast Project available at <http://www.dvb.org>
- [2] ISO/IEC 21000-1:2002, "Multimedia Framework-Part 1", 2002
- [3] ISO/IEC 21000-7:2004, "Information technology – Multimedia framework (MPEG-21) – part 7: Digital Item Adaptation", First Edition, Octobre 2004.
- [4] D. Kutscher, J. Ott, and C. Bormann, "Session Description and Capability Negotiation", IETF MMUSIC WG Internet-Draft, <http://www.ietf.org/internet-drafts/draft-ietf-mmusic-sdpng-08.txt>, Minneapolis, USA, February 2005.
- [5] X.sun, S.i, F.Wu, G.Shen, and W.Gao, "Efficient and flexible drift-free video bitstream switching at predictive frames", ICME 2002.
- [6] X.sun, F.Wu, S.Li, W.Gao, and Y.Q.Zhang, "Seamless Switching of Scalable Video Bitstreams for Efficient Streaming", IEEE International Symposium on Circuits and Systems, ISCAS 2002, Scottsdale, Arizona, USA, 26-29 May, 2002.
- [7] Y.K. Chou, L.C. Jian, and C. Wen Lin, "MPEG-4 video streaming with drift-compensated bitstream switching," in Proc. IEEE Pacific-Rim Conf. Multimedia, pp. 847-855, Dec. 2002.
- [8] S. Wee, J. Apostolopoulos and N. Feamster, "Field-to-Frame Transcoding with Temporal and Spatial Downsampling," IEEE International Conference on Image Processing, October 1999.
- [9] E. Amir, S. McCanne, and H. Zhang "An Application Level Video Gateway", In Proc. ACM Multimedia 1995.
- [10] E. Amir, S. McCanne, and R. Katz "An Active Service Framework and its application to real-time multimedia transcoding", in SIGCOMM, symposium on communications architectures and protocols, Septembre 1998.
- [11] M.Handley, S.Floyd, J.Padhye, and J.Widmer "RFC 3448, TCP Friendly Rate Control (TFRC): Protocol Specification" Request for Comments, IETF, Jan. 2003.
- [12] S.Floyd, M. Handley, J. Padhye, and J. Widmer "Equation-based congestion control for unicast applications", In Proc. of ACM SIGCOMM, pp. 43–56, 2000.
- [13] J. Rosenberg, and H. Schulzrinne "RFC 2733, RTP Payload Format for Generic Forward Error Correction", Request For Comments, IETF, Dec. 1999
- [14] I. Rhee and S. Joshi, "Error recovery for interactive video transmission over the Internet", IEEE J. Selected Area Comm., vol. 18, no. 6, pp. 1033-49, June 2000
- [15] S. Shenker, J. Wroclawski, "RFC: 2216 Network Element Service Specification Template", Request for Comments, IETF, sep. 1997.
- [16] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss "RFC: 2475 An Architecture for Differentiated Services", Request for Comments, IETF, December 1998.
- [17] T. Guenkova-Luy, A. Schorr, F. Hauck, A. J. Kassler, I. Wolf, B. Feiten, C. Timmerer, and W. Romijn, "Harmonization of Session and Capability Descriptions between SDPng and MPEG-21 Digital Item Adaptation", IETF MMUSIC WG Internet-Draft, Minneapolis, USA, February 2005.
- [18] T. Guenkova-Luy, A. Schorr, F. Hauck, A. Kassler, I. Wolf, B. Feiten, Ch. Timmerer, H. Hellwagner, "MPEG-21 DIA based content

delivery using SDPng controls and RTP transport" ISO/IEC JTC1/SC29/WG11, MPEG2004/M12002, Busan (Korea), April 2005

[19] VideoLan open source project available at <http://www.videolan.org/vlc/>

[20] Freecom web site <http://www.freecom.com>.

[21] H. Schulzrinne, A. Rao, R. Lanphier, "RFC: 2326 Real Time Streaming Protocol (RTSP)", Request for Comments, IETF, April 1998.

[22] FFMPEG open source project available at <http://ffmpeg.sourceforge.net/index.php>

[23] T.Ahmed and I.Djama "Delivering Audiovisual Content with MPEG-21 enabled Cross-Layer QoS Adaptation" accepted in the International Packet Video Workshop PV'06, China (Hangzhou), April 2006.