A MULTIMEDIA TERMINAL FOR ADAPTATION AND END-TO-END QOS CONTROL

Beilu Shao[1], Marco Mattavelli[1], Daniele Renzi[2], Maria T. Andrade[3], Stefano Battista[2],
Samuel Keller[1], Giorgiana Ciobanu[3], Pedro Carvalho[3]

[1]Swiss Federal Institute of Technology (EPFL), Switzerland,
[2]bSoft Ltd, Italy, [3]INESC Porto, Portugal
{beilu.shao, marco.mattavelli, samuel.keller}@epfl.ch, {daniele, bautz}@bsoft.info,
{maria.andrade, giorgiana.ciobanu, pedro.carvalho}@inescporto.pt

ABSTRACT

This paper addresses multimedia end user system design for content distribution over heterogeneous networks and terminals, with particular focus on End-to-End quality of service (QoS) control. A multimedia terminal comprising content-related metadata processor, usage environment characteristics provider, end user QoS monitor and H.264's extension Scalable Video Coding (SVC) audio-visual player in coordination under a terminal middleware, has been conceived and implemented. This end user terminal enables End-to-End QoS control for content adaptation solution both in semantic and physical approaches to maximize end user's perceptual experience and minimize resources. Such design approach illustrates a possible architecture for next generation multimedia end user terminal supporting MPEG-21 and H.264 extension SVC codec standards.

Index Terms— multimedia end user terminal, end-to-end quality of service, MPEG-21, H.264/SVC

1. INTRODUCTION

With the rapid development of multimedia computing and great success of the Internet, real-time multimedia applications have gained tremendous attention and contribution from both academia and industry. Indeed, different techniques have been developed to enable efficient and effective multimedia content delivery ranging from advanced video coding with excellent rate-distortion performance to adaptive multimedia delivery. Contemporarily, various transmission and terminal devices have been deployed, from PC, Set-Top Box to mobile devices like PDA. Also, a set of new content formats, standard-compliant or proprietary, have emerged resulting in a diversity of media itself. As a result, multimedia communication has been growing from single format video transmission over the Internet to complex multimedia adaptive delivery across heterogeneous networks and terminals [1].

In particular, the Internet has been growing from a monolithic data service network to next-generation network convergence of services. Content providers, service providers, network providers and other entities collaborate to enable value-added services for end users. As an example, real-time media services are enabled with End-to-End QoS guarantees by resources provisioning, monitoring and control. The combination of End-to-End QoS control and dynamic multimedia adaptation provides a promising approach to real-time multimedia applications [2].

However, current terminal design rarely considers the requirements imposed from the above. This paper addresses this new arisen problem, notably with the support for the end-user QoS and inter-operability for adaptation. We identify the following key mechanisms within a terminal architecture: 1) a mechanism to process content-related information, i.e. metadata or content descriptors, using MPEG-21 standard-based inter-operable format; 2) a mechanism to describe end user environment and user preferences at terminal, e.g., the processor capabilities of the terminal; 3) a mechanism to decode scalable audio-visual compressed bit streams, e.g., a media player with scalable video codec support for flexible adaptation in spatial, temporal, quality dimensions in real-time; 4) a mechanism to probe perceptual QoS for feedback, to trigger an Adaptation Decision-Taking Engine to reduce bit rate due to limited network bandwidth. These components achieve the goal of providing customers a service with the best possible quality given the constraints on terminals, networks and the cost of service.

The main contributions of the paper are: 1) design and implementation of a terminal consisting of the above identified components, supporting the combination of content adaptation and QoS monitoring system solution; 2) implementation of a terminal middleware to coordinate the components and optimization of the entire architecture with a distributed approach to reduce complexity; 3) support and exploitation of various aspects of MPEG-21 and MPEG SVC standards, providing an open standard-based solution.

The paper is organized as follows. Section 2 explains the terminal design of each component and highlights inter-operability for heterogeneous content delivery, End-to-End QoS control, SVC adaptation approach. Section 3 presents

the implementation details. Finally, Section 4 draws the conclusion and discusses future research perspectives.

2. TERMINAL ARCHITECTURE

2.1. Digital item browser

To obtain a terminal as interoperable as possible, the content should be represented also in an inter-operable format, more precisely based on the MPEG-21 Digital Item Declaration (DID) [3]. This has imposed the need to develop a module, the Digital Item Browser, to process the MPEG-21 DIDs and present them in a user friendly manner.

The MPEG-21 DID defines a standard XML-based model to express the structure of complex digital objects and to describe its components (media resources or additional descriptions). These media components do not need to have a specific encoding format, nor do the descriptions need to follow a given metadata model. The DID are represented in a standard XML-based language.

Though achieving interoperability across formats and systems, the user friendly presentation of MPEG-21 DIDs involves a significantly complex processing step. Consequently, a distributed approach is used to decouple processing tasks, time and power consuming, from rendering tasks. Two distinct modules, client and server, communicating via a well defined Web Services interface using Simple Object Access Protocol (SOAP) are designed.

Client-server architecture provides a clear separation between the Digital Item processing tasks and the operations related to its presentation to the user. The server application of Distributed DI Browser is named Digital Item Processing Server (DIP Server) and the client sub-system is generically designated as Graphic Digital Item Renderer (GDI Renderer), as shown in Fig. 1.
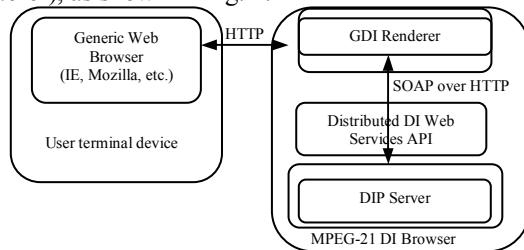

Fig. 1. Distributed DI Browser Architecture

2.2. QoS probes and QoS mapping

In order to analyze the perceptual quality at the terminal side and provide sufficient support for the adaptation modules, perceived Quality of Service (pQoS) probes have been developed at end user terminal. The measured values of perceived quality are periodically provided to the adaptation modules for adaptation-taking decision process.

During the process, a series of QoS mapping are calculated. Initially, an objective approximation of the perceptual quality is taken. Such an approximation is provided by objective QoS, which is extracted from objective quality parameters, like bit rate, frame rate, packet loss, etc. The objective QoS is mapped to the subjective QoS with specific mapping algorithms, which have been extensively addressed and reported. Then, the perceived QoS has to be mapped to the adaptation QoS (aQoS). Such a mapping allows a correspondence between perceived quality levels and practical adaptation parameters to make adaptation modules aware of the application class that should correspond to quality level measured at the terminal.
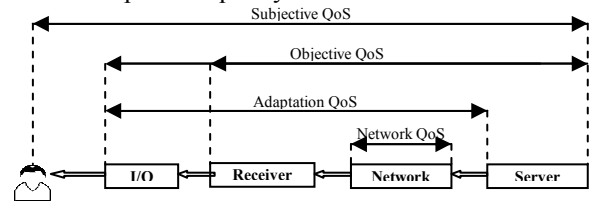

Fig. 2. Subjective, Objective, Adaptation, Network QoS

For adaptation, the quality information is provided in terms of absolute values of objective QoS, rather than alarms to leave the entire decision to adaptation modules. Therefore, the latter must be aware of the pQoS-aQoS mapping for each terminal. E.g., the same pQoS could correspond to a certain service level (e.g., Silver) for IPTV services and higher levels (e.g., Gold) for UMTS services.

1) Subjective QoS to Objective QoS Mapping: A subjective evaluation of the perceptual quality is difficult due to various factors including time, cost and human perception. Alternatively, objective quality assessment algorithms have been proposed. Objective QoS can be assessed by analyzing the signals in both compressed (e.g., MPEG-4 compressed video stream) and non-compressed (e.g., reconstructed RGB video) formats. A Mean Opinion Score (MOS) rating ranged under a standard scale is usually used for subjective tests in [4] and summarized in Table 1. The objective MOS is based on the same scale [5].

Table 1. Subjective MOS and R Factor scale

Subjective Interpretation	MOS	R
Excellent – almost all satisfied	5	90
Very good – most users satisfied	4	80
Good – some users unsatisfied	3	70
Fair – many users unsatisfied	2	60
Poor - most users unsatisfied	1	50

The R Factor is the rating factor used by the E-model [6], which is another rating system for conversational systems. E-model provides an objective measurement of quality, normally based on packet loss, jitter and delay as main quality parameters. We use for our 'multimedia system' a QoS model based on the same approach used by the E-model, because it provides a general and reasonable way to address the evaluation of objective quality.

2) Perceived QoS to Adaptation QoS Mapping: the adaptation decision-taking modules need to be provided with perceived QoS to adaptation QoS mapping to get a

correspondence between perceived QoS and the initial deployed service QoS. Adaptation QoS relates to the facilities embedded within an application that preserves the quality of its intended usage, by providing the levels of application class which correspond to the quality levels perceived by the user. For example, each frame rate (aQoS) corresponds to a perceived quality (pQoS) level. A performance threshold can be set against each application class. Such a mapping can be agreed between the service provider and probe constructors.

Usually three levels of aQoS are used (Bronze, Silver, Gold), any of which is mapped to a specific range of pQoS (MOS). For example, in a simple scenario performing frame rate adaptation by temporal scalability, the frame rate is the main perceived QoS parameter. The perceived frame rate must correspond to an SVC temporal layer in the source stream. That is why a mapping between pQoS and aQoS is needed, as example in Table 2.

Table 2. aQoS – pQoS mapping example

aQoS	Gold	Silver	Bronze
SVC Temporal Layer	2	1	0
Frame Rate[fps]	25.000	12.500	6.250
Target pQoS [R factor, scale 0 – 100]	85	70	55
pQoS Interval [R factor, scale 0 – 100]	10	10	10
pQoS Range [R factor, scale 0 – 100]	80 - 90	65 – 75	50 – 60
MOS [scale 1 – 5]	4.4	3.8	3.2

3) Network and Media QoS Parameters: The objective QoS parameters monitored can be grouped into two domains: a) Network quality parameters, like packet loss, jitter, etc; b) Media (Audio and Video) quality parameters, like bit rate, audio saturation, video frame size, etc.

2.3. SVC player and adaptation approach

The recent advances in video coding led to new standards H.264/MPEG-4 Advanced Video Coding (AVC) [7] and its extension Scalable Video Coding standards (SVC) [8]. AVC is the first one that has achieved a significant improvement on video coding efficiency though at high computation expenses and SVC enables a single stream with multiple temporal, spatial, quality scalabilities, while keeping compression at a high efficiency. The focus of this section is adaptation approach based on SVC.

The achievement of different scalabilities co-existing in a single bit stream allows video adaptation at bit stream level from SVC. Such benefits increase adaptation flexibility. For example, if a terminal is limited by certain constraints and its decoder supports SVC, there is no need for intermediate adaptation, since the receiver can perform the adaptation itself by discarding the relevant Network Abstraction Layer (NAL) Units that convey enhancing layers. This is particularly useful for mobile devices.

In other cases, mainly due to network constrains, like in wireless environments, an upstream adaptation is necessary,

e.g., to reduce bit rate to lighten the network load. In such scenarios, if the adapted bit stream already fulfils the terminal needs, the terminal do not need to drop enhancing scalable layers, as a network adapter will do it. Such adaptation can be either transcoding or translation. In the former approach, if the stream is provided to a specific type of terminals, with the same constraints of computing and bandwidth, adaptation modules can perform transcoding from SVC to AVC by removing unnecessary scalability header information to save some bandwidth. In the latter approach, the stream can be merely cut to keep necessary layers in the remaining sub-streams, which form an Operation Point (a set of layers necessary to reconstruct a given quality in temporal, spatial, quality dimensions). These sub-streams can be produced and transmitted in a layered multicast mode, which is with different RTP sessions (different IP address and UDP port) carrying different scalable layers, to serve various kinds of terminal with different streams.

In a framework with end-to-end QoS guarantees, as the one in this paper, an upstream adaptation can be directly demanded by the terminal to a Media Aware Network Element (MANE), performing adaptation operations. The terminal can request the adaptation by means of feedback messages, carrying QoS metrics, which can be subsequently processed by MANE for adaptation decisions. The MANE relies on the scalability features of the source SVC stream to provide the quality level demanded by the terminal. Thus, scalable video coding facilitates an elegant and flexible adaptation framework, serving a variety of terminals and networks with the same video content.
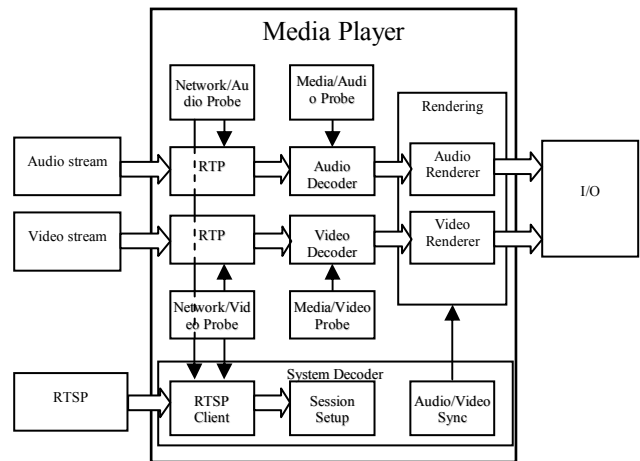
Fig. 3. Media Player Architecture

The video decoder included in the Media Player implements complete functionalities to decompress digital video sequences from binary streams according to either one of the supported coding standards (MPEG-4 Video, H.264 /AVC and its extension SVC). To achieve this goal, the decoder is organized for a complete reuse of the low level functionality of a video decoder. In particular, the

AVC/SVC Decoder supports a wide range of picture formats, picture rates and bit rates. Correspondingly, it supports various types of content scalability and, within any of these, multiple layers, to provide the maximum level of adaptation. The Media Player architecture is designed to easily integrate also other terminal modules, especially the pQoS probes which are plugged into relevant points of the Media Player (e.g., RTP Receiver, RTSP parser, Audio and Video Decoders) to get QoS information concerning various levels of the media content as shown in Fig. 3.

2.4. UED-UCD module

User Environment Description (UED) includes descriptive information related to user characteristics, (e.g., user information and user preferences), terminal capabilities (e.g., codec capabilities and display capabilities), network characteristics (e.g., available bandwidth, delay, and error), and natural environment characteristics (e.g., location and time). User Constraints Description (UCD) supplements UED to provide further constrain. These are typical tools used for digital content adaptation.

2.5. Terminal middleware

The components are all coordinated under a distributed terminal device middleware (TDM). In principal, such middleware has two responsibilities: 1) externally interacting with server or intermediate service, e.g., to feedback for QoS control for content adaptation or to contact with DIBrowser@Server for content browsing; and 2) locally coordinating the convergence of DIBrowser@Client, Media Player, UED-UCD, QoS Probes.
.

3. TERMINAL IMPLEMENTATION

This section presents the implementation details. To have a focus, we only abstract the most functional part and highlight them in a block schema as shown in Fig. 4. The entire multimedia terminal has been designed on Windows, Windows Mobile and Symbian platforms in order to maximize the portability to various platforms. Such implementation exploits the requirements of heterogeneity and inter-operability. As a result, it allows the terminal to be easily installed on multiple devices, such as TV, Set-Top-Boxes, PCs, PDAs and mobile phones.

4. CONCLUSION AND FUTURE WORK

The End-to-End QoS control for multimedia content delivery over heterogeneous networks has been investigated from multimedia terminal perspective. We propose a new end user terminal architecture including MPEG-21 Digital Item Browser, usage environment characteristics provider, end user's QoS monitor and Scalable Video Coding (SVC) audio-visual player coordinated under a terminal

middleware. Several adaptation methods have been proposed and the system architecture also has been optimized. Such a terminal facilitates content adaptation solution to maximize user's fruitful experiences of video quality. The architecture design also exploits and supports the MPEG-21 and H.264 extension SVC codec standards. In the future, we will introduce key management for intellectual property management and protection (IPMP).
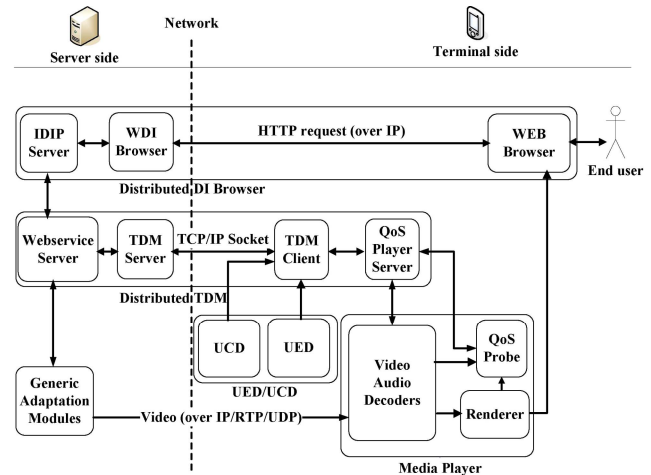

Fig. 4. Terminal Architecture

5. REFERENCES

[1] A. Vetro, C. Christopoulos, and T. Ebrahimi, "Special Issue on Universal Multimedia Access", *IEEE Signal Processing Magazine*, vol. 20 (2), Mar. 2003.
[2] T. Ahmed, A. Asgari, A. Mehaoua, E. Borcoci, L. B. Equille, and K. Georgios, "End-to-end quality of service provisioning through an integrated management system for multimedia content delivery", *Computer Communications*, vol. 30 (3), 2007, 638-651.
[3] ISO/IEC TR 21000-2:2003 "Information Technology - Multimedia Framework (MPEG-21)-Part2: Digital Item Declaration", 2003.
[4] T. Alpert, V. Baroncini, D. Choi, L. Contin, R. Koenen, F. Pereira, H. Peterson, "Subjective Evaluation of MPEG-4 Video Codec Proposal: Methodological Approach and Test Procedures", *Signal Processing: Image Communication,* vol. 9, 1997, 305-325.
[5] ITU-T P.800.1, "Mean Opinion Score (MOS) terminology", 2006.
[6] ITU-T G.107, "The E-model, a computational model for use in transmission planning", 2003.
[7] G. Sullivan, T. Wiegand, D. Marpe, and A. Luthra. "Advanced video coding for generic audiovisual services", ITU-T Recommendation H.264 and ISO/IEC 14496-10, ITU-T and ISO/IEC, 2005.
[8] T. Wiegand, G. Sullivan, J. Reichel, H.Schwarz, and M. Wien. "Scalable video coding", ISO/IEC 14496-10 Amendment 3, ISO/IEC and ITU-T, Jul. 2003.