Signal Processing: *Image Communication* 27 (2012) 192–207

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Secure transport and adaptation of MC-EZBC video utilizing H.264-based transport protocols ☆

Hermann Hellwagner [b], Heinz Hofbauer [a,*], Robert Kuschnig [b], Thomas Stütz [a], Andreas Uhl [a]

[a] *Department of Computer Sciences, University of Salzburg, Jakob-Haringer-Straße 2, 5020 Salzburg, Austria*
[b] *Institute of Information Technology, Klagenfurt University, Universitätsstraße 65–67, 9020 Klagenfurt, Austria*

ARTICLE INFO

Article history:
Received 6 May 2011
Accepted 15 November 2011
Available online 22 November 2011

Keywords:
Scalable Video Coding (MC-EZBC)
In-network adaptation
RTP/SRTP MANE
Video encryption
Selective encryption
Format compliance

ABSTRACT

Universal Multimedia Access (UMA) calls for solutions where content is created once and subsequently adapted to given requirements. With regard to UMA and scalability, which is required often due to a wide variety of end clients, the best suited codecs are wavelet based (like the MC-EZBC) due to their inherent high number of scaling options. However, most transport technologies for delivering videos to end clients are targeted toward the H.264/AVC standard or, if scalability is required, the H.264/SVC. In this paper we will introduce a mapping of the MC-EZBC bitstream to existing H.264/SVC based streaming and scaling protocols. This enables the use of highly scalable wavelet based codecs on the one hand and the utilization of already existing network technologies without accruing high implementation costs on the other hand. Furthermore, we will evaluate different scaling options in order to choose the best option for given requirements. Additionally, we will evaluate different encryption options based on transport and bitstream encryption for use cases where digital rights management is required.

1. Introduction

The use of digital video in today's world is ubiquitous. Content consumers desire to retrieve content through a multitude of networks, from 3G to broadband Internet, on a broad range of consumer devices, from cell phones to high performance PCs. However, consumers do not care about the technicality necessary to provide the content over this wide range of networks but rather about their quality of experience (QoE), i.e., they want to consume the best possible quality in a timely manner. This creates a problem for content providers since it is costly, in both time and storage space consumption, to provide content for every conceivable end device and network link. Re-encoding on the other hand is expensive in the way that it requires significant time which reduces the QoE for end users.

The solution to this problem is called Universal Multimedia Access (UMA) [1]. The goal of UMA is to encode content once and adapt it in a timely manner to current end user requirements. One of the enabling technologies of UMA is the use of scalable video coding. This averts the need for transcoding on the server side and enables the server to scale the video. However, even scaling requires computation time and reduces the number of connections the server can accept. Furthermore, variable bandwidth conditions, which happen frequently on mobile devices, further tax the server with the need to adapt the video stream. The solution to this is usually in-network adaptation, shifting the need to scale to the node in the network where a change in bandwidth is occurring. The core

☆ Supported by Austrian Science Fund (FWF) project P19159-N13.
* Corresponding author.
E-mail addresses: hermann.hellwagner@uni-klu.ac.at (H. Hellwagner), hhofbaue@cosy.sbg.ac.at (H. Hofbauer), robert.kuschnig@uni-klu.ac.at (R. Kuschnig), tstuetz@cosy.sbg.ac.at (T. Stütz), uhl@cosy.sbg.ac.at (A. Uhl).

0923-5965/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2011.11.002

adaptation with these restrictions takes place on the server and adaptation due to actual channel capability is done in-network.

For video streaming in the UMA environment, i.e., a high number of possible bandwidths and target resolutions, wavelet based codecs should be considered. Wavelet based codecs are naturally highly scalable and rate adaptation as well as spatial and temporal scaling is easily achieved. Furthermore, wavelet based codecs achieve a coding performance similar to H.264/SVC, cf. Lima et al. [2]. Under similar considerations Eeckhaut et al. [3] developed a complete server to client video delivery chain for scalable wavelet-based video. However, there are already standardized ways of transporting multimedia data, namely the Real-time Transport Protocol (RTP) [4]. Similarly, there is a protocol for handling a single or several time-synchronized stream of continuous media, e.g., audio and video, the Real Time Streaming Protocol (RTSP) [5] which can use RTP as its mode of transportation. Besides RTP and RTSP the MPEG-21 Part 7 'Digital Item Adaptation' (DIA) [6] can be used to provide content related metadata. A codec agnostic description, the generic Bitstream Syntax Description (gBSD) [7], can also be used as a basis for an informed adaptation process.

In order to use existing technology, i.e., RTP streaming and in network adaptation, modules for handling the motion compensated embedded zero bit codec (MC-EZBC) have to be created to facilitate packetization for RTP and media awareness for adaptation nodes. However, the existing technology can already deal with H.264/SVC, e.g., [8] describes the H.264/AVC payload for RTP and multimedia aware network elements (MANE) and [9] extends this to H.264/SVC. Since the H.264/* bitstream is build from network abstraction layer units (NALUs), the fastest route to utilize the existing infrastructure is to encapsulate the MC-EZBC into a NALU bitstream which presents itself as H.264/SVC to those components. Following this route it is, apart from the MC-EZBC to NALU conversion, trivial to use the existing infrastructure. Also note that, while we only take a look at MC-EZBC to NALU conversion, such a conversion can be constructed for other scalable video codecs and the theoretical and experimental analysis will by and large also hold for those conversions.

In this paper we will provide a method of encapsulating the MC-EZBC into a NALU bitstream. Additionally, we will investigate how this encapsulated bitstream can be transported, encrypted, and scaled, and at what cost in terms of payload overhead and network delay. Furthermore, we will look at surrounding issues which have to be taken into account, e.g., initial vectors for encryption.

In Section 1.1 we will describe the basics of the chosen wavelet based video codec, the MC-EZBC, in Section 2.1 a description of the layout of the bitstream will be given and the adaptation to the RTP packetization scheme will be given in Section 2.2. An overview of the MPEG-21 DIA generic Bitstream Syntax Description (gBSD) will be given in Section 2.3. Section 2.4 describes additional requirements for the RTP streaming process for the MC-EZBC and presents the outline of the encapsulation process.

The main concern of research regarding UMA is usually performance with respect to scaling and in-network adaptation. However, digital rights management and security are also a prime concern for providers of commercial videos. Furthermore there are a range of other aspects of video streaming, ranging from server requirements to protocols, to QoS, etc., Wu et al. [10] give a good overview of these aspects. General principles and possible goals of digital rights management (DRM) will be explained in Section 1.2 and application of encryption to the MC-EZBC codec will be discussed in Section 3.

In Section 4 we will compare the different aspects and options of the adaptation and streaming process theoretically and experimentally.

1.1. The motion compensated embedded zero bit codec (MC-EZBC)

For reasons of scalability which fit the UMA principle we use the enhanced MC-EZBC wavelet based video codec for in-network adaptation. This choice was made mainly because the source code is available,[1] which enables our experiments. The MC-EZBC codec [11–14] is a scalable t-2D video codec which uses motion compensated temporal filtering, with 5/3 CDF wavelets, followed by regular spatial filtering, with 9/7 CDF filtering, an overview of the encoding pipeline is given in Fig. 1a. This method, temporal first and spatial later, is referred to as t+2D coding scheme, see Fig. 1b for an example of this decomposition for a group of picture (GOP) size of 8. For temporal filtering a full decomposition is used and thus the GOP size is discernible by the number of temporal decomposition levels. Both temporal and spatial filtering is done in a regular pyramidal fashion. Statistical dependencies are exploited by using a bit plane encoder, the name giving embedded zero bit coder. Motion vectors are encoded with DPCM followed by an arithmetic coding scheme.

For an overview of wavelet based video codecs and a performance analysis as well as techniques used in those codecs see the overview paper by Adami et al. [15]. Again, while we concentrate only on the MC-EZBC in this paper the encapsulation process described later can in a modified version still be applied to other scalable video codecs. Likewise the analysis performed will also be indicative for other scalable video codecs.

1.2. Overview of encryption and digital rights management

Shannon's work [16] on security and communication shows that the highest security is reached through a secure cipher operating on almost redundancy free plain text. Current video codecs exploit redundancy for compression and we can consider the bitstream to be a redundancy free plain text in the sense of Shannon. Thus for maximum security we just need to encrypt the whole bitstream with an state of the art cipher, i.e., the Advanced Encryption Standard (AES) [17]. However, the choice was made to keep information in plain text in order to facilitate scalability in the encrypted sequence. Regarding security, Lookabaugh

[1] The source for the ENH-MC-EZBC is available from http://www.cipr.rpi.edu/research/mcezbc/.