Cross-Layer Optimized Transmission of H.264/SVC streams over DVB-T2 Broadcast System

Lukasz Kondrad, Imed Bouazizi, Vinod K. M. Vadakital, Miska M. Hannuksela, Moncef Gabbouj

***Abstract*—DVB-T2 has been recently standardized by the DVB organization.**

In this paper the deployment of H.264/SVC in DVB-T2 is discussed. We focus on the system configuration aspects which provide optimal use of DVB-T2 features from the H.264/SVC point of view. We also propose modifications to the error protection mechanisms that would consequently lead to an improvement of user experience. The data dependencies in H.264/SVC are exploited to protect the media data unevenly according to their priorities.

***Index Terms*—Broadcasting, Scheduling, Algorithm**

I. Introduction

As analogue switch-off (ASO) approaches in a number of countries, and digital television is steadily gaining a large interest from users, the DVB organization decided to design a new physical layer for digital terrestrial television. The main goals of the new standard were to achieve more bandwidth compared to DVB-T, targeting HDTV services, improve single frequency networks (SFN), provide service specific robustness, and target services for fixed and portable receivers. As the result of work carried by the DVB organization, the DVB-T2 specification [3] was released for publication on the DVB website in June 2008. Initial tests shows that the new standard achieves more than 40% capacity improvement compared to DVB-T.

Parallel to DVB's work on DVB-T2 standard, the Joint video team (JVT) group worked on an extension to H.264/AVC which defines scalable video coding (H.264/SVC)[1]. The addendum is published as an annex in the existing H.264/AVC standard, and was finalized in November 2007. H.264/SVC bit-stream is constructed such that it contains one base layer and one or more enhancements layers. The base layer is fully backward compatible with the H.264/AVC bit-stream, while each enhancement layer can improve the video in temporal, spatial, and/or quality (SNR) domain. H.264 SVC extension was developed as an alternative to the simulcast solution and exploits the redundancy between different representation (quality, spatial, and temporal) layers. Simulations [4] show better savings in bandwidth when using H.264 SVC in comparison to Simulcast. Recently, the newly developed SVC extension has attracted good interest from industry and research organizations alike. For instance, H.264 SVC is being adopted by the DVB organization as one of the video codecs used for DVB broadcast services [2].

The existing standards for broadcast applications are not optimized for the transmission of scalable video streams and need to be tailored to exploit all the benefits of the H.264 SVC codec. In our previous work [5], the deployment of H.264 SVC in DVB-H was discussed. We focused on the system configuration on the link layer. However, transmitting scalable signal can also be achieved by using hierarchical modulation [6].

The new DVB-T2 standard aimed at providing service specific robustness. Therefore, it could be a good transmission system for delivery of scalable video bitstream. However, the standard introduces limitations which prevent it from delivering one service over more than one physical layer pipe. Due to this limitation, the layer specific robustness of scalable service cannot be implemented on the physical layer. We propose a cross-layer optimized scheduling method used in the input pre-processor sub-system of DVB-T2. The proposed method uses information from physical and application layers to avoid fragmentation of the more important media data units. By consequence, less important media data units will endure the penalty of fragmentation induced packet losses. Hence, the proposed approach results in unequal error resilience for scalable video streams.

The rest of this paper is organized as follows. In the next Section, background information about DVB-T2, H.264/SVC and use cases of H.264/SVC over DVB-T2 transmission are provided. Then, different options for multiplexing/ scheduling of H.264/SVC streams are presented in Section III. Simulation results are presented and discussed in Section IV. Finally, the paper is concluded in Section V.

This work was supported by the Academy of Finland, (application number 129657, Finnish Programme for Centres of Excellence in Research 2006-2011)

Lukasz Kondrad, Vinod K. M. Vadakital, and Moncef Gabbouj are with the Signal Processing Department, Tampere University of Technology, Finland, FI-33720 (lukasz.kondrad@tut.fi, vinod.malamalvadakital@tut.fi, moncef.gabbouj@tut.fi).

Imed Bouazizi and Miska M. Hannuksela are with Nokia Research Center, Finland, FI-33720 (imed.bouazizi@nokia.com, miska.hannuksela@nokia.com).

II. Background

This section provides background information needed and is organized in the following way. First, details of the DVB-T2 features are described. Next, the main concept of scalable video codec is provided. At the end H.264/SVC over DVB-T2 transmission's use cases are described.

A. DVB-T2 Broadcast System

DVB-T2 standard specifies the physical layer structure by defines the construction of the over-the-air signal which is produced by a T2 modulator. The processing at the receiver side is left open to different implementation solutions.

The input to the T2 system consists of one or more logical data streams where one logical data stream is carried by one Physical Layer Pipe (PLP). For video services this would include video/audio streams plus associated signalling (for example PSI/SI information). PLP is a fully transparent data pipe which generally enables transporting data of whatever structure with freely selectable, but PLP-specific physicals parameters. Due to this feature, not only capacity, but also the service robustness can be adjusted to particular needs, depending on the type of terminal and its usage environment. On Figure 1 the high level architecture of the DVB-T2 system is shown.

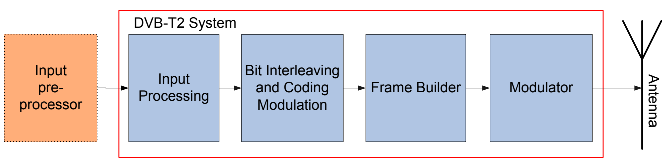

Fig. 1 High Level Architecture of DVB-T2 system

The input pre-processor module is not part of the T2 system but is a logical unit which may be included and it could work as a service splitter, scheduler or de-multiplexer.

The input processing module operates separately on the contents of each PLPs. It slices the input data stream into data fields and inserts a Base Band (BB) header at the start of each data field which consequently form a BB frame. The size of the BB frame is fixed for a given PLP and it depends on the forward error correction code rate which is later applied on the BB frame. Two types of frames are provided. The first is short with the frame size varying between 3072 and 13152 bits and the second is normal, with the size varying between 32208 and 53840 bits, see Table 1.

Table 1 BB frame size in bits

LDPC	Kbch	
Code	Short frame	Long frame
1/2	7 032	32 208
3/5	9 552	38 688
2/3	10 632	43 040
3/4	11 712	48 408
4/5	12 432	51 648
5/6	13 152	53 840

In the Bit Interleaving and Coding Modulation block, each BB frame is processed and a FEC frame is generated. Outer coding (BCH) followed by inner coding (LDPC) and then bit interleaving is performed on each BB frame. The parity check bits of the systematic BCH outer code are appended after the BB frame data field, and the parity check bits of the inner LDPC encoder are appended after the BCH field, as shown in Fig. 2. The FEC frame has fixed size, 16200 bits for short frame and 64800 bits for normal frame.

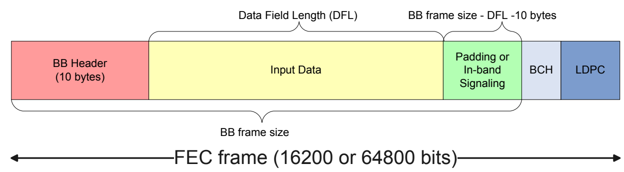

Fig. 2 FEC frame structure

The resulting FEC frames are passed to a Frame Builder module, where interleaving and mapping to physical layer frames, as well as OFDM symbol mapping is performed. The resulting physical layer frames are then passed to the Modulator module for modulation and transmission.

B. Scalable Video Coding

Scalable video coding paradigm has been widely investigated in academia and industry for the last 20 years. However, before H.264/SVC standard scalable video coding was always linked to increased complexity and drop in coding efficiency when compared to non-scalable video coding. Hence, scalable video coding was rarely used and it was preferred to deploy simulcast, which provides similar functionalities as a scalable video coding bit stream by transmission of two or more single layer streams at the same time. Though simulcast causes significant increases in resulting total bit rate, there is no boost in the complexity.

The new H.264/SVC standard is an extension to H.264/AVC standard which allows for temporal, spatial and quality scalability in a video bit-stream. However, contrary to the previous implementations of scalable video coding, H.264/SVC is characterized by a good coding efficiency and moderate increase in complexity, and therefore, it can be seen as a reasonable alternative to the simulcast.

The idea behind scalable video coding is that encoder produces a single bit-stream containing different representations of the same content with different scalable qualities. A scalable video coding decoder can then decode a subset of the bit-stream that is most suitable for its use case and capabilities. A scalable bit stream consists of a base layer and one or more enhancement layers. The removal of enhancement layers leads to a decoded video sequence with reduced frame rate, picture resolution or picture fidelity. The base layer of H.264/SVC is an H.264/AVC bit-stream which ensures backwards compatibility to existing H.264/AVC receivers. Through the use of scalable video coding, improved spatial resolution adaptation, bit rate adaptation, and/or even power adaptation can be achieved. One main use case for H.264/SVC lies in the exploitation of the intrinsic media data importance, (e.g. based on the scalable video coding layer those media units belongs to) to achieve better error resiliency using unequal error protection. Enhanced service consumers (those consuming the base and enhancement layers) may then benefit from the graceful degradation achievable in case of packet losses or transmission errors.

When temporal scalability is used, frames from higher layers can be discarded, which results in a lower frame rate, but does not introduce any distortion during play out of the video. In the case of spatial scalability, the encoded bit-stream contains sub-streams that represent the same content at different spatial resolutions. Quality scalability enables the achievement of different operation points, each yielding a different video quality. Coarse Grain Scalability (CGS) is a form of quality scalability that uses the same tools as the spatial scalability. Medium Grain Scalability (MGS) achieves different quality encoding by operating on the transform coefficients.

C. H.264/SVC over DVB-T2 transmission

Service specific robustness is one of the tools, which comes with DVB-T2 standard. This feature combined along with information from the dependent nature of scalable video bitstream, may be utilized to improve robustness of video transmission. For example, a H.264/SVC encoded service may be split into base layer and enhancement layer streams by an IP router or input pre-processor module. Consequently base layer stream could be transmitted by PLP with higher protection while enhancement layer would be carried by less robust PLP which subsequently would results in unequal error protection of the transmitted data. The example of that setting is presented on Figure 2, below.

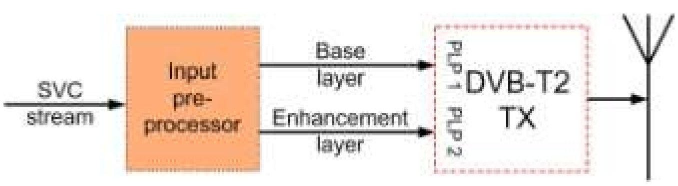

Fig. 3 SVC transmission over DVB-T2

That setup is similar to transmission of SVC over DVB-H using hierarchical modulation, which has already been shown to bring benefits [6]. However, in the DVB-T2 specification it is assumed that a receiver will always be able to receive one data PLP and its associated common PLP, if any [3]. That statement does not exclude the approach presented on Fig. 3 but does not assure that the service transmitted in foregoing scenario will be compatible with the legacy terminals.

Therefore, H.264/SVC bit stream should be transmitted over one PLP, unless not specified otherwise. However, the hierarchical structure of SVC and the priority information it brings, may be utilized to bring more robust transmission in other manners, which will be presented in the next Section.

III. Cross-Layer Optimized Scheduling for DVB-T2

Transmission errors after physical decoder are reflected on the BB frame level. It is assumed that if the combined BCH/LDPC FEC decoding fails, then the whole BB frame is discarded and marked as lost. In this paper, we propose a scheduling algorithm for optimized mapping of service data to the data fields of the BB frames. The scheduler constitutes a component of the pre-processor at the DVB-T2 transmission chain. One scheduler is allocated per each PLP, in order to operate on the data packets of that PLP. The pre-processor and the contained components are depicted in Figure 4.

The proposed scheduling algorithm avoids fragmentation of the IP packets that contain media data of higher importance. By avoiding fragmentation of important media units, improved error resilience is achieved. A fragmented media unit will generally be discarded if one or more of its fragments are corrupted.

Additionally, restricted time interleaving is applied to IP packets that contain media units of a higher importance access unit. Time interleaving spreads the media units of an access unit across multiple T2 frames. By consequence, losses which are typically of a bursty nature, would most likely not affect the complete access unit. As an example, an IDR picture that consists of several slices would ultimately be mapped into several BB frames that are spread over multiple T2 frames. Transmission errors may corrupt a set of consecutive BB frames depending on the burst length. Due to the time interleaving, the impact of loss of a set of consecutive BB frames would less likely result in significant loss to the random access points.

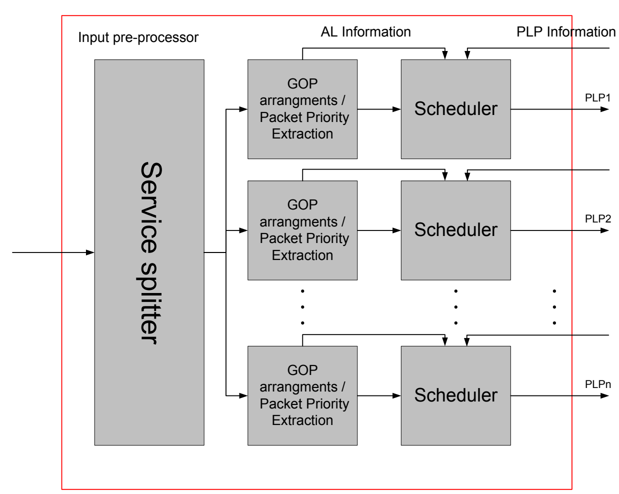

Fig. 4 Cross layer optimized input pre-processor concept

As mentioned beforehand, the time interleaving is restricted to limit the required initial buffering time and to keep the channel switch time within an acceptable range. The number of T2 frames that are used for the time interleaving of the random access point and the related group of pictures is restricted to 1 to 1.5 seconds. With a typical T2 frame duration of 250ms, the total number of T2 frames used for time interleaving a group of pictures is then 4 to 6 T2 frames.

The size of the data field in a BB frame for a specific service depends on the selected modulation scheme and physical layer FEC code rate. Upon determining the size of the payload of BB frame, the number of BB frames needed to transmit the set of pictures of the video stream can be calculated based on the total size of the media units to be transmitted. The number **M** of BB frames allocated for the service in each T2 frame can be dynamically determined according to the following equation:

$$M = \frac{S}{\text{Payload Size of BB frames} \times N}$$

M: Number of BB frames/T2 Frame allocated for service

N: Number of T2 frames

S: Total Size of Media Units over duration of N T2 Frames

After determining the BB frame allocation over the set of T2 frames, the scheduling algorithm proceeds by mapping media data packets to BB frames. The target thereby is manifold. First, the mapping algorithm avoids fragmentation of important media units over more than one BB frame. Secondly, it aims at providing maximum error resilience through time interleaving. Finally, the algorithm aims at increasing bandwidth usage efficiency by avoiding total fragmentation overhead and padding operations.

The above discussed problem is similar to the bin packing problem and is an NP-hard problem. We follow a heuristic solution to keep the complexity within a manageable range while achieving a close to optimal solution. The algorithm is described in the sequel:

1) Arrange media packets in descending order of importance
2) Start from higher importance media packets (e.g. those containing base layer IDR pictures) and assign them to maximally distant BB frames.
3) For the rest of the media packets, order media packets according to their size in decreasing order
4) Loop though the set of media packets and
   - 4.1) Assign packet to the best fitting BB frame (the BB frame that leaves the least free space after adding the media packet)
   - 4.2) If no fitting BB frame is found queue the media packet at the tail of the set of media packets
   - 4.3) Stop if no media packet can be mapped to available free space
   - 4.4) End Loop
5) Fragment the left-over media packets starting from the first BB frame

This proposed scheduling algorithm is best suitable for scalable media such as an SVC media stream. The scheduler complexity is limited to the handling of the RTP packet header and the RTP payload format. Given that the set of media encoding options in a broadcast scenario is limited, this additional functionality would not significantly increase the complexity of the scheduler.

IV. Evaluation

In this section, we present and discuss the results of a comparison of scheduling method presented in Section III and generic approach without scheduling.

For the simulation Crew and Crowd sequences, both with 1280x720 resolution and 600, 500 frames long respectively, were used. The sequences were encoded using main profile of H.264/AVC encoder. To create a simple temporal scalability non-reference b frames every second reference frame was employed. This meant that a base layer with 15 fps, and enhancement layer with 30 fps were created. The encoding parameters were set as follows: bitrate ~8 Mbits/s, IDR frame was put every 30 frames, and the maximum slice size was set to 1300 bytes. H.264/AVC encoder was used instead of H.264/SVC to allow better error concealment algorithms at the decoder side.

To conduct the simulation an Input Pre-Processor (IPP) module presented on Fig. 4 and describe in Section III was implemented. To simulate the physical transmission over DVB-T2 bearer, a BB frame error pattern generated by DVB-T2 physical layer simulator was utilized. Four different error patterns were used throughout simulations, each one with different BB error rate and different data field length of BB frames. In the error pattern, a BB frame was marked as lost when the BCH/LDPC decoding failed.

To properly reproduce the input of the Input Pre-Processor module to each NAL units of the encoded sequence an additional 67 bytes long header was added which corresponded to GSE/IP/UDP/RTP headers. The scheduler sub-module of IPP module was working on packet belonging to one GoP (30 frames).

At the receiver side the error of the BB frames were mapped on the data packets and if any of a fragmented part of the packet was marked as lost then whole packet was marked as lost. Next, the lost NAL units were discarded from the error free sequence and erroneous bitstream was decoded using H.264/AVC decoder with motion vector copy error concealment method.

The following configurations of the scheduler have been analyzed in the simulations:

1) A generic approach without scheduling. The scheduler based on data field length of BB frame fragments the packets as they come and adds new GSE header and CRC check to fragmented packet.
2) A cross-layer approach where scheduler uses information from the physical layer (data filed length of BB frame) and application layer (priority of the packet). Based on that information an algorithm, described in Section III, was examined.

In Table 2 and Table 3 PSNR value and packet loss rate are depicted for each of the tested configurations for Crew and Crowd sequences respectively. It can be seen that thanks to cross layer scheduling approach the packet loss rate can be reduce and in consequence around 0.5 dB PSNR gain achieved.

Table 2 PSNR and Packet error rate (Crew)

BB frame error rate [%]	Generic	Proposed Scheduling Algorithm	
7.33	9.12	9.03	Packet error rate [%]
	28.72	29.19	PSNR [dB]
3.32	4.42	4.43	Packet error rate [%]
	32.48	32.74	PSNR [dB]
1.80	2.22	2.14	Packet error rate [%]
	34.93	35.63	PSNR [dB]
1.55	2.03	2.01	Packet error rate [%]
	35.26	35.66	PSNR [dB]
0.00	39.85	39.85	PSNR [dB]

Table 3 PSNR and Packet error rate (Crowd)

BB frame error rate [%]	Generic	Proposed Scheduling Algorithm	
7.33	8.81	8.67	Packet error rate [%]
	23.81	24.04	PSNR [dB]
3.32	4.48	4.45	Packet error rate [%]
	26.15	26.00	PSNR [dB]
1.80	2.23	2.18	Packet error rate [%]
	27.48	27.99	PSNR [dB]
1.55	2.07	2.04	Packet error rate [%]
	27.95	28.09	PSNR [dB]
0.00	30.88	30.88	PSNR [dB]

The gain in PSNR is achieved not only by packet loss reduction but as well thanks to spreading errors through less important packets. In Table 4 and Table 5, below, the number of error packets together with number of error packets due to fragmentation is presented.

Table 4 Number of error packets (Crew)

BB frame error rate [%]	Generic			Proposed Scheduling Algorithm			
	I	P	B	I	P	B	
7.33%	169	1034	236	139	913	373	error packets (ep)
	72	419	91	0	246	342	ep due to fragmentation
3.32%	88	481	129	54	470	175	error packets (ep)
	43	229	67	0	154	174	ep due to fragmentation
1.80%	38	255	58	36	204	97	error packets (ep)
	13	94	20	0	27	95	ep due to fragmentation
1.55%	33	219	68	21	184	112	error packets (ep)
	16	92	32	0	39	101	ep due to fragmentation

Table 5 Number of error packets (Crowd)

BB frame error rate [%]	Generic			Proposed Scheduling Algorithm			
	I	P	B	I	P	B	
7.33%	142	801	210	137	707	291	error packets (ep)
	57	321	82	0	183	287	ep due to fragmentation
3.32%	93	427	66	75	368	140	error packets (ep)
	47	204	32	0	165	121	ep due to fragmentation
1.80%	27	226	39	33	161	91	error packets (ep)
	10	85	13	0	26	81	ep due to fragmentation
1.55%	31	201	39	34	141	92	error packets (ep)
	14	88	18	0	17	90	ep due to fragmentation

The results show that due to scheduling none of the packets belonging to I frame is lost because of fragmentation. Additionally, due to applied time interleaving to the "I" packets less of them is affected by errors. Moreover, it is proven that the proposed scheduling methods move most of the errors to the packets belonging to the less important B frames.

On Fig. 4 PSNR plot for the first 60 frames is presented. It can be seen that thanks to the error packet spread the rapid quality change can be avoided (frames 10 – 30).

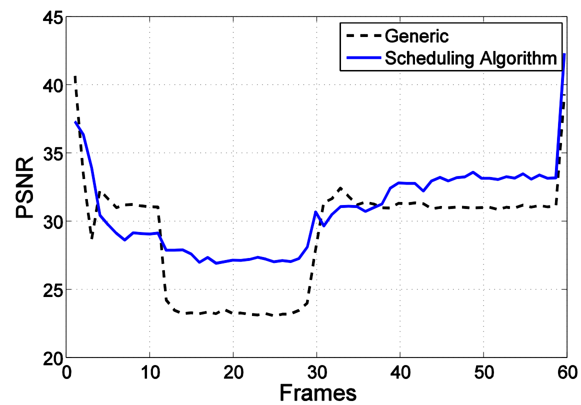

Fig. 5 PSNR of Crew sequence for first 60 frames

V. Conclusion

In this paper, a scheduling algorithm for optimized mapping of service data to the data fields of the BB frames was proposed and examined through simulations. The results proved that thanks to the scheduling algorithm the errors are moved to the less important data packets which lead to unequal error resilience of the transmitted stream. In spite of the PSNR results does not show significant gain when the scheduling is used, we think that with decoder employed with better error resilience tools the proposed scheduling will be valuable input to the DVB-T2 transmission system.

References

[1] ITU-T Recommendation, "H.264: Advance Video Coding for generic audiovisual services", November 2007.

[2] ETSI TS 102 005 V1.4.1, "Digital Video Broadcasting (DVB): Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols", Working Draft.

[3] ETSI EN 302 755 V1.1.1, "Digital Video Broadcasting (DVB):Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)", Draft July 2008.

[4] M. Wien, H. Schwarz, and T. Oelbaum "Performance Analysis of SVC", IEEE Transition Circuits System. Video Technology., vol. 17, no. 9, pp. 1194–1203, September 2007.

[5] I. Bouazizi, L. Kondrad, M. M. Hannuksela, Moncef Gabbouj "Efficient FEC Protection of Scalable Media Streams in DVB-H", IEEE 58th Annual Broadcast Symposium, Virginia USA, 15 - 17 October 2008

[6] C. Hellge, T. Schierl, and T. Wiegand,"DVB-H with hierarchical modulation using SVC", www.dvb.org TM-AVC 0362, March 2008.