H2B2VS (HEVC HYBRID BROADCAST BROADBAND VIDEO SERVICES) – BUILDING INNOVATIVE SOLUTIONS OVER HYBRID NETWORKS

R. Monnier[1], A. Mourelle[2], J-P. Bernoux[3]
C. Alberti[4], J. Le Feuvre[5], W. Hamidouche[6]

[1]Thomson Video Networks, France; [2]Hispasat, Spain; [3]TDF, France
[4]Ecole Polytechnique Fédérale de Lausanne, Switzerland
[5]Institut Mines-Télécom, France; [6]IETR/INSA, France

ABSTRACT

Broadcast and broadband networks continue to be separate worlds in the video consumption business. Some initiatives such as HbbTV have built a bridge between both worlds, but its application is almost limited to providing links over the broadcast channel to content providers' applications such as Catch-up TV services. When it comes to reality, the user is using either one network or the other.

H2B2VS is a Celtic-Plus project aiming at exploiting the potential of real hybrid networks by implementing efficient synchronization mechanisms and using new video coding standard such as High Efficiency Video Coding (HEVC). The goal is to develop successful hybrid network solutions that enable value added services with an optimum bandwidth usage in each network and with clear commercial applications. An example of the potential of this approach is the transmission of Ultra-HD TV by sending the main content over the broadcast channel and the required complementary information over the broadband network. This technology can also be used to improve the life of handicapped persons: Deaf people receive through the broadband network a sign language translation of a programme sent over the broadcast channel; the TV set then displays this translation in an inset window.

One of the most important contributions of the project is developing and testing synchronization methods between two different networks that offer unequal qualities of service with significant differences in delay and jitter.

In this paper, the main technological project contributions are described, including SHVC, the scalable extension of HEVC and a special focus on the synchronization solution adopted by MPEG and DVB. The paper also presents some of the implemented practical use cases, such as the sign language translation described above, and their performance results so as to evaluate the commercial application of this type of solution.

INTRODUCTION

Started in January 2013, H2B2VS (1) is a Eureka Celtic-Plus project. It is aiming at investigating the hybrid distribution of TV programs and services over broadcast and broadband networks. The technology used for the contents compression is the emerging video coding standard: High Efficient Video Codec (HEVC).

Coordinated by Thomson Video Networks, the consortium is composed of 21 partners.

The project is organized around four main Work Packages (WP). The goal of the first one is to study uses cases and business models to guarantee that the work done by the project is in line with the market's expectations. WP2 is dealing with the technologies to be adapted in order to allow HEVC hybrid distribution. All these technologies are integrated in WP3 to build demonstrators which are used by WP4 in public demonstrations as a basis to disseminate project's results throughout the industry. One of the major outcomes of WP4 is also the contribution to the relevant standardization bodies in order to guarantee a successful commercial deployment of the technology.

MAIN TECHNOLOGICAL CHALLENGES

Among all the challenges the project had to face, two are predominant for the use cases which are described in this paper. The first one is scalable coding and the second one is synchronization.

Scalable Video Coding (SHVC)

The Scalable High efficiency Video Coding (SHVC) is the scalable extension (2) of the HEVC standard. SHVC extension (3) was finalized in July 2014 by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) under a partnership known as Joint Collaborative Team on Video Coding (JCT-VC).

SHVC defines tools to provide spatial, fidelity, bit depth, and color scalability. The SHVC syntax elements, mostly signalled at the Video Parameter Set (VPS) header, are common to all HEVC extensions. These syntax elements provide information on the video layers such as resolution, bit depth and inter-layer dependencies. The SHVC encoder consists of L HEVC encoders, one encoder to encode each layer with L the number of layers. The first layer is the Base Layer (BL), followed by L-1 Enhancement Layers. In the case of SHVC spatial scalability, the Base Layer HEVC encoder encodes a down-sampled version of the original video and feeds the first Enhanced Layer encoder with the decoded picture and its Motion Vectors (MV). The enhancement layer encoder l (l=2,…, L) encodes a higher resolution video, using the decoded picture from lower layer as an additional reference picture. The inter-layer reference picture is up-sampled and its MVs are up-scaled to match the resolution of the layer being decoded. Figure 1 shows an example of an SHVC encoder encoding two layers in spatial scalability configuration. In the case of SNR scalability, the encoding process remains unchanged except that the picture used for inter-layer prediction is used without being up-sampled and its MVs up-scaled. As shown in Figure 1, the outputs from the two encoders are multiplexed to form a conforming SHVC bit stream.
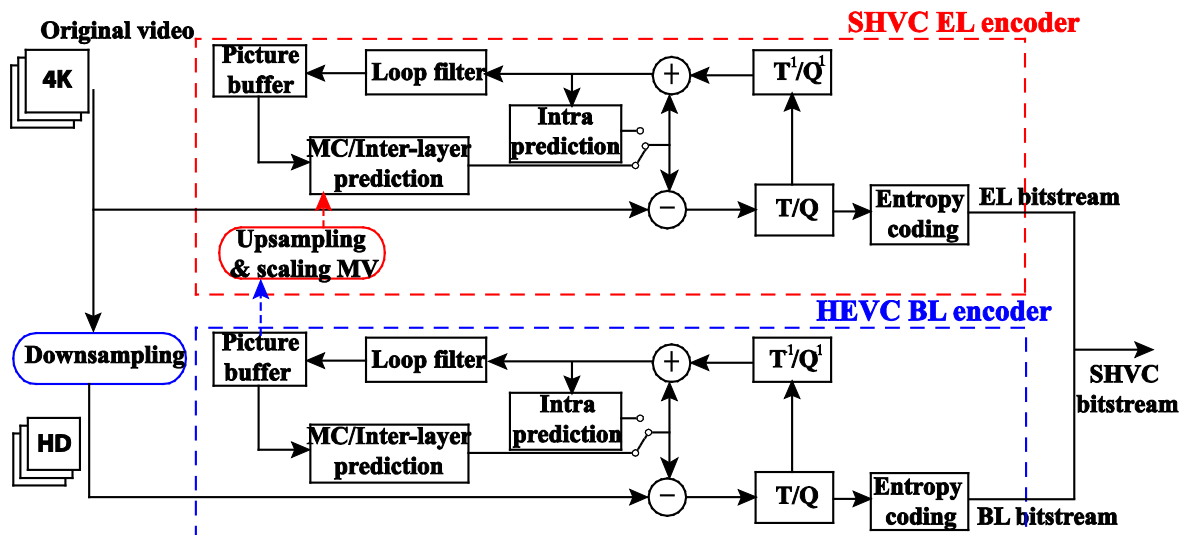
Figure 1 - Block diagram of an SHVC encoder (two layers, 2x spatial scalability)

Synchronization

Combining media streams originating from both broadcast and broadband networks impacts the end-to-end chain in various ways: The most important ones are synchronization and buffer management. Synchronization of media streams may depend on the scenario: A second screen advertisement may not require more than one second accuracy, while a picture in picture of an alternate view may require around or less than 100 ms. In the most demanding cases such as scalable video coding, synchronization has to be frame accurate, since media from both streams must be re-aggregated before entering the decoder.

Most if not all deployed broadcast TV networks rely on the MPEG-2 transport stream to carry media data. The Transport Stream has been designed at a time where clock precision of TVs and set-top boxes was not reliable enough to ensure proper audio-visual buffer management and synchronization of the content and, therefore, uses its own clock, called Program Clock Reference (PCR), to instruct the receiving entity when data shall be removed from input demultiplexer or decoder buffers and when decoded data shall be presented. The PCR and the time stamps associated with each audio or video frames is not related to the original media time of the content, may not be known by the service packaging the broadband media streams and may be modified by network equipment such as transcoders. This is why it cannot be used in a reliable way to synchronize the broadcast and broadband networks for video services. In order to identify a common, precise time origin for both broadband and broadcast streams, several approaches have been used: Audio fingerprinting or audio watermarking tools are commonly used in existing second-screen application, but they require additional processing of the media, implying higher energy consumption. They are furthermore not precise enough to achieve frame-accurate synchronisation for layer-coded video. A simpler approach is to embed a timeline in the broadcast stream, different from the PCR and carried along by the various equipment of the broadcast chain. This timeline is usually a media time code or a network time such as NTP. The frequency at which this timeline is sent usually depends on the

synchronization requirement, and may be sent once per video frame for use cases requiring a high-precision.

The second issue faced in hybrid delivery scenario is buffer management: The broadcast network has a constant latency while broadband networks suffer from potentially high jitters and congestion. In order to reduce the effects of congestion, several solutions have been developed to deliver media streams over either UDP or TCP, most of them relying on bitrate adaptation of the media. The adaptation can be done at the sender side, as it is typically the case with RTP streams in VoD or video-conferencing, or at the client side, as used in video over HTTP (HLS, Smooth Streaming, MPEG-DASH). The bandwidth adaptation relies on either alternate encoding of the media, or scalable layers of coded data that can be thinned out of the media stream. Handling of the jitter requires a reception buffer large enough to smooth latency variations. When synchronizing and de-jittering a broadcast-broadband setup, the broadcast has to be delayed long enough for the de-jitter buffer to work. This can be achieved by buffering the broadcast video at the client side, but this impact an extra start-up delay of the hybrid media session. Alternatively, the broadcast stream can be buffered before the modulation stage so that broadband streams may be requested before the corresponding broadcast data arrives at the receiver.

In H2B2VS, the following design choices were made regarding these issues:
- The synchronization is achieved by inserting in each broadcast video frame a media time code.
- Most of the buffering needed for the broadband data is given by the delay induced by the DVB-T2 transmission chain, which is around 2 seconds.
- The adaptive broadband network is using MPEG-DASH protocol: The media segments are encoded at the same time as the broadcast channel, using a segment length (GOP size) of one or two seconds. When using two seconds, the de-jittering at the client side relies entirely on the client buffer. For 1s, the DASH client is able to fetch media segment 1s before their presentation is due, allowing for a proper de-jittering.

H2B2VS partners did not stop at developing a technology and test it. They were also very active in standardization: They proposed the so-called "TEMI" time line to MPEG, accepted in MPEG-2 TS and also adopted by DVB and HbbTV.

MPEG started its activity towards the standardization of a synchronization mechanism of auxiliary media streams for its Systems technologies (MPEG-2 Transport Stream, ISO Base Media File Format, MPEG-DASH, MPEG Media Transport) at the 102nd MPEG meeting in 2012. During this meeting, a solution was proposed by the H2B2VS partners to enable carriage of DASH timeline values in MPEG-2 TS in order to unambiguously recover media time information, regardless of MPEG-2 TS time discontinuities, with a low-overhead of typically 7 kbps for a 60Hz video stream, compared to 90kbps of existing similar solutions. In the following two years, the H2B2VS partners proposed several contributions refining the solution, with a strong focus on frame accuracy synchronization. This activity led to an amendment to the MPEG-2 Systems standard for the introduction of a new descriptor for the carriage of Timeline and External Media Information (TEMI).

Figure 2 shows how this descriptor is carried in a MPEG-2 TS as a "pointer" to external media carried by a different channel (e.g. broadband). The synchronization among the two streams is implemented via the insertion of TEMI descriptors carried in adaptation field of broadcast Transport Stream packets. A TEMI descriptor essentially contains timestamps

and location descriptors providing the URLs of the currently running associated extension, or the upcoming ones.
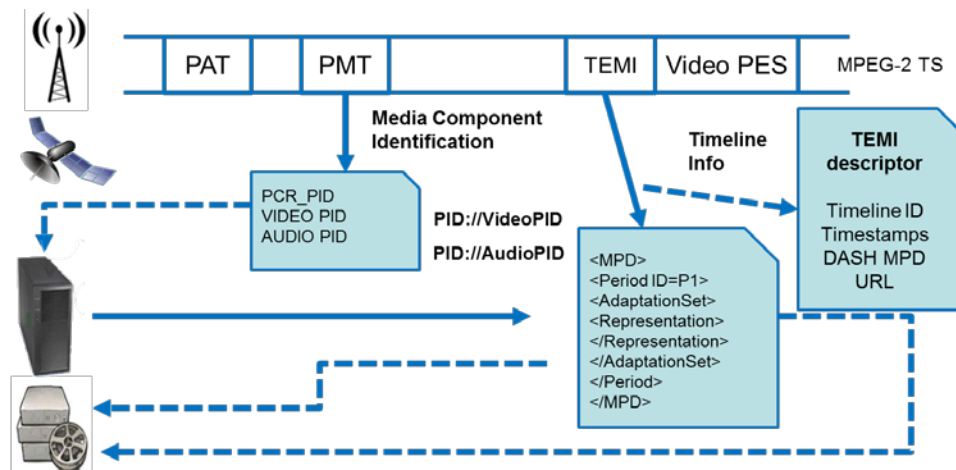

Figure 2 –TEMI descriptor in the MPEG-2 Transport Stream

During the work described above, MPEG issued several Liaison statements with DVB in order to harmonize the synchronization mechanisms used by the two standardization bodies. The TEMI proposal was positively received and supported by DVB. After the final publication of the MPEG-2 Systems amendment, the initial synchronization mechanism where the timeline is carried in the private data bytes of the adaptation field was superseded by the use of the TEMI descriptor.

Likewise, the HbbTV consortium followed the work done by MPEG and the H2B2VS partners on the TEMI timeline descriptor and version 2.0 of the HbbTV standard proposes it as a mean to implement multi-stream synchronization. The section devoted to APIs for media synchronization provides an example of synchronization of a broadcast channel with a DASH stream delivered over broadband relying on the use of TEMI descriptor.

USE CASES

Overview Of The Uses Cases Studied By H2B2VS

During the first part of the project, the H2B2VS partners defined twenty use cases where the use of hybrid and HEVC technologies could lead to new services. The result was the definition of four categories of use cases:

- Picture quality improvement: Basically, the broadcast network carries a standard High Definition (HD) programme and additional data is sent over the broadband network to allow the display of an enhanced picture.

  Several enhancements are proposed. 4K is the most obvious one. However, studies done by the project are targeting also High Frame Rate (HFR), High Dynamic Range (HDR) and Wide Gamut. Where relevant, scalable approaches are considered.

- Customized TV: Here, the broadband network is used to increase the experience of the end user watching a basic broadcast programme.

Trick modes (see below) or additional audio services – quality improvement or multi-language - are some examples. Programme personalization is also envisaged: The end user would be able, for instance, to choose the end of a drama among several scenarios or to select in a Formula 1 race the car the picture comes from. Personalized advertising conveyed by the broadband network and inserted in broadcast programs is another interesting use case because it would allow targeting the advertising as it is done today on the Web.

- <u>Picture in Picture & 2nd screen</u>: Several use cases are dealing with the display of alternate/additional contents conveyed by the broadband network, either on the main screen or on the companion screen.

  The end user can follow-up another programme coming through the broadband network and displayed in a "Picture In Picture" (PIP) way, without the need for a second tuner. The PIP picture can also be the sign language translation of the main programme coming through the broadcast network.

- <u>Network improvements</u>: Using the broadband network to relieve the broadcast network is also possible with the technology developed by H2B2VS.

  For example, when a severe outage occurs on a satellite reception, the broadband network can take over to ensure an uninterrupted TV service. To cope with the problem of spectrum scarcity on terrestrial networks, it can be envisaged to off-load the terrestrial network by transmitting programmes with a low audience on the broadband network, on the basis of dynamic audience measurements. Regional variations of a national programme carried by the broadband network can also be a good solution to optimize the use of the satellite or terrestrial spectrum.

The twenty use cases were not implemented by the H2B2VS project. Based on business criteria, partners defined a subset of use cases and prototypes were developed. Some of these Uses Cases are detailed below.

Sign Language Translation

Today, the penetration of sign language video services is slowed down because everyone, hearing-impaired or not, is forced to watch them. This drawback comes from the way the sign language video translation is inserted in the main programme: It is done in the control room. With H2B2VS' solution, the broadcast programme remains unchanged: The broadband network is used to carry the sign language translation. Only hearing-impaired people will watch it and the PIP insertion is be done by the TV set at end-user's request. Of course, the synchronization mechanism is key to ensure that the translation is aligned with the main programme.

The benefit is twofold. For the end-user as explained above and for the TV channel who does not have to pay for additional bit rate over the broadcast network to convey the sign language translation on an extra PID. The solution proposed by the H2B2VS project is especially interesting in countries where sign language translation is required by law during a certain amount of time: It allows TV channels to fulfil their duties at a reduced cost.

Hybrid Distribution Of Scalable 4K

This is a way to offer a 4K (Ultra High Definition) experience to end-users having connected TVs and watching HD TV programmes on broadcast networks. SHVC is used

to send on the broadcast network an HD programme and an enhancement layer on the broadband network. The terminal uses these two layers to decode and display a 4K video, thanks to the picture-accurate synchronization mechanism.

This approach is an answer to introduce 4K services on terrestrial networks which suffer from spectrum scarcity. Further to the high pressure from mobile operators to release frequencies in the UHF band, in some countries, it seems difficult to introduce 4K services immediately without modifying substantially the way multiplexes are managed. The solution proposed by H2B2VS allows to carry-on broadcasting HD programmes and to offer a 4K enhancement through Internet. This is a way to optimize the bandwidth both on the broadcast and broadband networks.

Compared to a solution where the full 4K programme would be sent over Internet, bit rate is saved, increasing thus the number of end-users who can receive 4K programmes. Using figures of ADSL coverage available for the UK, we can estimate how the number of households receiving 4K can be increased thanks to hybrid distribution.

Let's assume the following for the distribution of a scalable 4K programme: The base layer is broadcast at 3.5 Mbps on the terrestrial network and the enhancement layer is sent at 8 Mbps over the broadband network. To get the same video quality, the bit rate over Internet should be 11 Mbps to carry a non-scalable 4K programme.

Figure 3 shows the distribution of broadband speeds in the UK (4). Depending on the population density, the increase of the number of households receiving 4K with a hybrid technology is different. In overall, the percentage increases from 58% to 70%. In Urban areas, the improvement is slightly smaller because bit rates such as 11 Mbps are quite usual: The reach goes from 68% to 78%. A significant higher gain is obtained in rural areas where the number of subscribers who would be able to receive 4K is doubled, going from 20% to 40%.

The results may vary from one country to another but this study gives a first idea of the potential of hybrid distribution technology to introduce 4K services where spectrum scarcity may be an obstacle.
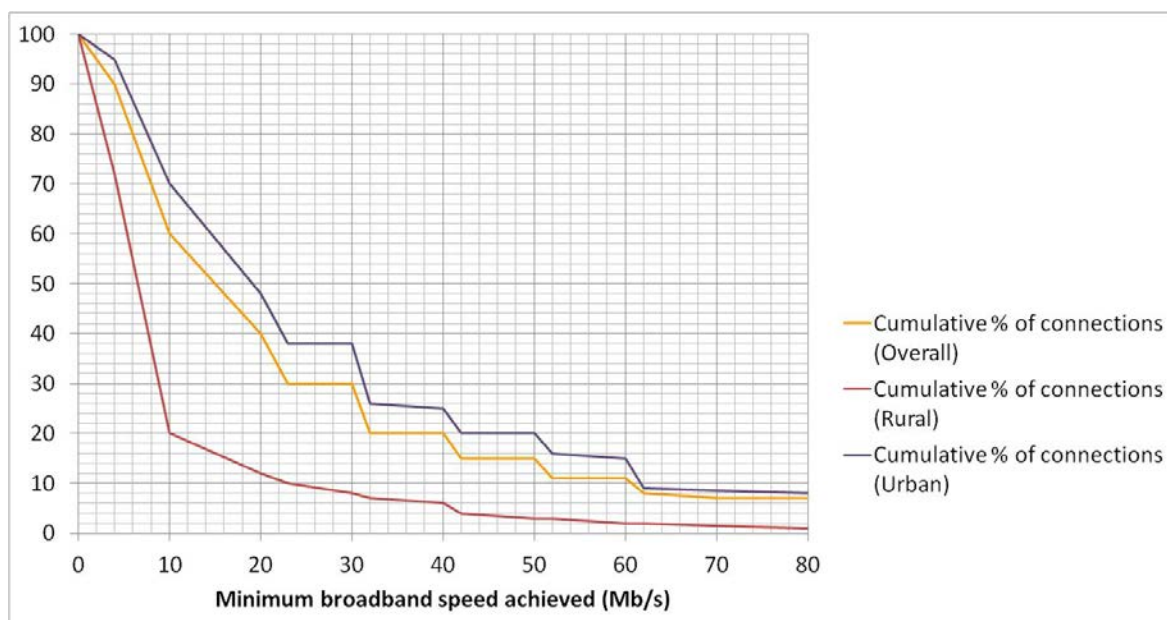

Figure 3 – Distribution of broadband speeds, by population density

Trick Mode (TDF)

When defining the use cases, H2B2VS's partners took into consideration the flexibility hybrid receivers can bring to TV consumption. Indeed, it is very important that linear TV, mostly represented by "broadcast" television, follows viewers' new practices. One of them is the "TV when I want" concept. The end-user must be able to withhold a TV programme and resume it on the same device or on another one, when available, without missing any part of the programme. The technology developed by H2B2VS allows it with no external Hard Disk Drive: Everything is in the network.

This is possible because the receiver is able to switch from broadcast reception to broadband reception. H2B2VS takes advantage of the possibility of the above-mentioned TEMI timeline. The TV programme is encoded simultaneously for both networks (broadcast and broadband) and the MPEG2 Transport Stream (MPEG2-TS) includes the TEMI information. To ensure the quality of the streaming on the broadband network, DASH adaptive streaming is used. The TEMI information is used to detect and store the moment when a broadcast programme was stopped and to identify the relevant DASH segment when the receivers has to switch to broadband when resuming the programme. The synchronization mechanism allows getting the relevant segment for a quasi-seamless switch. Tick mode keys such as "Fast Forward", "Rewind" are also available to navigate in the streams.

CONCLUSIONS AND FUTURE WORK

H2B2VS is close to its end (November 2015). It paved the way to offer new hybrid video services. Several demonstrators were developed, allowing the demonstration of several use cases. It is now time to get a feedback from the different actors of the ecosystem, including the final users. Partners will thus spend some time demonstrating these use cases in order to get this feedback, measure their business value and, hopefully, confirm the market studies done at the beginning of the project. Based on that, some developments may be carried out outside the project to turn out the H2B2VS prototypes into products.

REFERENCES

1. http://h2b2vs.epfl.ch/
2. G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, and A. Vetro "Standardized Extensions of High Efficiency Video Coding (HEVC)", IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 1001-1016, December 2013.
3. J. Chen, J. Boyce, Y. Ye, M. Hannuksela, G. J. Sullivan, Y. K. Wang HEVC Scalable Extensions (SHVC) Draft Text 7, JCTVCR1008, Sapporo, JP, July 2014.
4. Ofcom, Ofcom's second full analysis of the UK's communications infrastructures. Infrastructure Report 2014, 8 December 2014, page 41.

ACKNOWLEDGEMENTS