Video Distribution by Special Hardware Tools

More events now need 4K video for projection in high quality and with low latency

Sven Ubik, Jiri Navratil, Jiri Halak
Research department for network applications
CESNET a.l.o.
Prague, Czech Republic Country
ubik@cesnet.cz, jiri@cesnet.cz, halak@cesnet.cz

Jiri Melnikov
CESNET / Faculty of Information Technology
CTU at Prague
Prague, Czech Republic
Jiri.Melnikov@fit.cvut.cz

***Abstract*— This paper describes the technical options and the use cases of the 4K Gateway technology for specialized distributions of various forms of HD, 4K and 3D videos for high quality projections on large screens. 4K Gateway can be used for transmissions of medical, cultural or sports events or for scientific collaboration. It is based on an FPGA design, which adds minimal internal latency (from 3 ms). It allows to transmit up to eight video channels in both directions. Based on available bandwidth, it can work in an uncompressed or compressed mode. For compression it uses a JPEG 2000 encoding and decoding core.**

***Keywords*— HD Video; 4K video; 3D technology; cyber performance; live surgery.**

I. Introduction

Today, downloading a YouTube video usually takes just seconds. Videos are available in several quality levels, therefore almost any type of network can be used. But there is a significant difference if we want to see a video recorded for fun by a simple web camera or if we want to observe a video during complicated surgeries done by endoscopic tools, via microscopes or via the da Vinci Surgical System with special 3D cameras located inside the patient body. It is different when we transmit a cultural event in real time to a remote site where several hundreds of people are in a big hall. In such cases, the video resolution and the capacity of the network are critical parameters. For medical applications the highest possible resolution is desirable HD (High Definition) or 4K (4096 x 2160) formats. Also if we want to project the video with wide angle projectors on walls several meters wide and high, the highest resolution of 4K or 8K is needed otherwise the picture will be blurred.

For live transmissions for real-time collaboration, we need to minimize the end-to-end delay. Therefore, we use uncompressed transmissions, which require significant network bandwidth. To calculate the required bitrate for the transmission we should take into account the video resolution, frames per second and other parameters such as color encoding and possible compression. The resulting bandwidth for one full HD (1920 x 1080) channel can range from several Mb/s to approx. 1.5 Gb/s for the uncompressed transmission.

II. Transmitting technology

For the transmission of real time video from the source place (surgery theatre, concert hall, etc.) into the conference venue or into the theatre we need an end-to-end connection. A shared Internet or dedicated links can be used. Sometimes it is a combination of both methods. Particularly, the last segments to venues usually need to be upgraded by a dedicated connection installed for the event.

Video transmission can be done in several ways. Videoconferencing tools or streaming technology (e.g., using RTMP – Real Time Message Protocol) will deliver the video to anybody who will connect to an MCU (Multipoint Control Unit) or a streaming server. Current commercial videoconferencing devices use mostly the H.323 protocol family and heavy compression so that they can work over inexpensive links with minimal bandwidth. The required bandwidth for an HD channel is in the range from 128 Kb/s to 6 Mb/s, depending on the compression level. In our experience, 4 Mb/s is the lowest bandwidth for acceptable quality for medical transmissions. We have used them for a couple of years for transmissions of ophthalmological surgeries.

For Internet streaming, we usually use hardware H.264 encoders (e.g., Makito) to contribute the picture to a remote streaming server (such as Wowza). A web page then allows multiple users to individually connect using one of several distribution formats (FLASH video, HTML5) and resolutions. These encoders are small, very easy to use and compatible with common streaming servers. On the other hand, they introduce a very long latency, 5-10 seconds are not uncommon. Therefore, they are not suitable for side channels for the interaction with the surgeon.

One of the first uncompressed tools was DVTS (Digital Video Transport System) [1,4] developed in Japan. DVTS accepts image sources (e.g., digital camcorders, surgical instruments) over the IEEE 1394 interface and streams them over an IP network. The quality of the picture was mostly defined by the digital camera and lighting conditions. The system was heavily used in medicine. For example in 2011 we broadcasted via DVTSplus a neurosurgery executed by prof. Takanori Fukushima during his visit at Masaryk Hospital in Usti nad Labem to many Asian partners, see Figure 1. The necessary bandwidth for this system was 30-35

 ISBN: 978-1-61208-404-6

Mb/s per channel including audio. This was only a minor problem, especially in an academic environment where plenty of bandwidth is available.

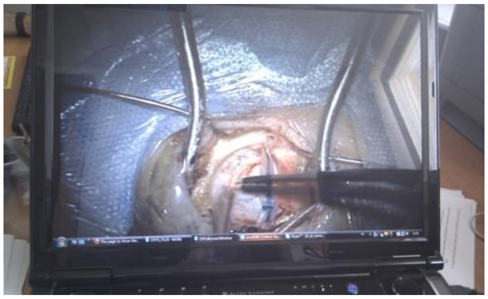

Figure 1. Neurosurgery image via DVTS

III. 4K GATEWAY

For low-latency real time transmissions of multi-channel video signals CESNET has developed an FPGA-accelerated solution called 4K Gateway [5]. The interface to the video input and output is through HD/3G-SDI interfaces and the interface to the network is by an interchangeable optical XFP transceiver. All video processing is implemented inside an FPGA (Field-Programmable Gate Array) that is by a specialized programmable hardware. This allows to eliminate all varying delays a PC platform causes by data copying, interrupts, device drivers, etc. The video processing is done by a sequence of firmware modules. Communication with firmware is through registers, which can be accessed from within an embedded Linux operating system, which runs on a Microblaze soft-core processor inside the FPGA.

Firmware can operate in uncompressed or JPEG 2000 modes. In the uncompressed mode, it can send and receive up to eight HD signals simultaneously, or up to two 4K signals. The processing latency on the sending and receiving device combined is as low as 3 ms. For higher resiliency to network jitter, additional buffering can be configured, at the expense of added latency. In the JPEG 2000 mode, the firmware can send or receive up to four HD signals simultaneously, or one 4K signal. The processing latency is currently 5 frames on the sending and receiving devices combined. Subframe latency based on multiple tile compression will be available in the future. The bitrate needed to transmit one HD channel depends on the frame rate, color encoding, whether embedded audio and ancillary data are transmitted and the level of compression. For example, for 1080p60 4:2:2 10-bit, the uncompressed video bit rate without packet overhead is 1920x1080 pixels x 60 frames x 2 color samples x 10 bits = 2.48 Gb/s. With JPEG 2000 compression, the bitrate for one HD channel is reduced to 20-100 Mbps depending on the quality requirements, with the higher end providing sufficient quality even for the most critical applications. Selected outputs at the receiving side can be synchronized at one-pixel precision, which allows high-quality 3D projections. This functionality allows to combine several resources (different cameras and digital video devices) and to create environment for perfect illusion of presence in a surgery theatre, concert stage or in the sport stadium.

An important property of the 4K Gateway device is the ability to maintain the sender - receiver synchronization without GPS or a similar timebase signal and keeping low latency. There are two sources of rate difference between the data arriving to the receiver from the network and the data that needs to be sent to the rendering device.

First, the internal clock of the sender can be different from the internal clock of the receiver, within a tolerance permitted by the respective transmission protocol formats. For example, the HD-SDI clock rate is specified as 1.485 Gb/s or 1.485/1.001 Gb/s with 100 ppm tolerance, while the 3G-SDI clock is twice this frequency.

Second, jitter can be introduced due to network traffic conditions. This network jitter needs to be accommodated by the receiver FIFO (First In First Out) memory.

Several alternative methods can be used to compensate the data rate difference: receiver feedback, frame buffer, blank period adjustments or rendering clock adjustments.

The receiver can send feedback to the sender requesting sending rate adjustments. This technique is used in window-based transport layer protocols, such as TCP or in some link layer protocols, such as PAUSE frames in Ethernet.

This can be too slow reaction and may require a large receiver FIFO memory, which would introduce high added latency. However, the main problem with this technique is that it requires a cooperating sender. The technique would not work with current cameras and other real-time video sources.

Frame buffer requires the receiver to have a FIFO memory large enough to accommodate several complete frames. Then the rendering device can be driven by a fixed clock oscillator in the receiver. After certain time when the skew between the sender and receiver clock rates causes the frame buffer to overflow of underflow, a frame skipping or duplication is used. In the worst case when both the sender and receiver clocks are shifted by 100 ppm in the opposite directions, the error can expand to the whole frame in 1/200 * 10^{-6} = 5000 frames. At 60 frames per seconds it is just 1.5 minutes. A large FIFO memory can extend this time at the cost of added latency.

A precise external clock source can be used to guarantee that the sender and receiver clocks are in sync. GPS receivers are commonly used for this purpose. A limitation of this method is that it is often difficult to get the GPS signal through the building to the sender or receiver location.

We use a method of rendering clock adjustments [7]. Adjusting the clock in HD-SDI channels between the receiver and the rendering device within the permitted tolerance gives the receiver some level of adaptation to the rate of incoming data. This solution requires tunable oscillators and a closed-loop controller in the receiver. In order to adjust the rendering clock to the data source rate, we used a common PID controller (proportional-integral-

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-404-6

derivative controler). The complete receiver control structure is shown in Fig. 2.

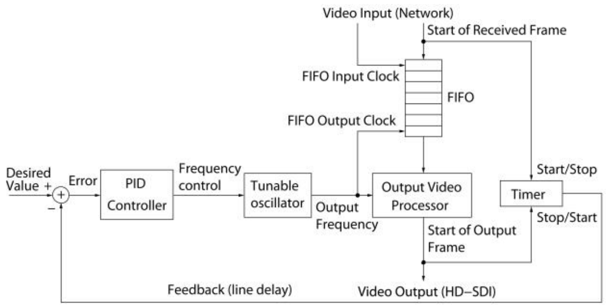

Figure 2. Rendering clock adjustment

The FIFO input is driven by data arriving from the network. The FIFO output is driven by a clock generator, which is tuned by the PID controller. The FIFO occupancy is used as a controlled variable. The desired value is set for required jitter accommodation and latency acceptance. The feedback value is taken from the current FIFO occupancy, which is sampled by the regulator every 200 ms. The controller then uses a weighted moving average of eight last samples. The purpose of this average is to smooth out fluctuations in FIFO occupancy due to network jitter. The PID controller then produces adjustments to the clock generator frequency.

Pairs of channels can be used for 3D signals and multiple channels can be used for tiles of beyond-high-definition signals. A frame detector looks for the beginning of frames in each specified group. After synchronization is applied, the frame generator is temporarily stopped until all channels in a group are aligned.

The device includes multiple tunable oscillators along with their own control loops as in Fig. 2. This allows to flexibly configure all SDI outputs into groups, where each group is independently adjusting its speed to the sender. In this way, one device with eight SDI outputs can receive and present several multi-channel signals in parallel, for example, one 4K signal, one 3D HD signal and two independent HD signals.

IV. Transmission Examples

We have provided multiple telepresence events, with applications mostly from the medical and cultural fields. The events were focused on demonstrations of new possibilities when using fast international research networks, such as GLIF (Global Lambda Integrated Facility) [2,3].

In 2010, CESNET started intensive collaboration with the Robotic Center of Masaryk Hospital in Usti nad Labem. This is one of the few hospitals in Czech Republic equipped with the robotic daVinci Surgical System, which uses a 3D camera to provide the surgeon a stereoscopic vision of the surgery area.

Figure 3. 3D medical transmission from the daVinci Surgical System

In Fig. 3 we show an example of a 3D medical transmission [6] from Usti nad Labem to the hospital in Banska Bystrica in Slovakia. We used this technology for transmissions also to far destinations, to the KEK research institute in Tsukuba, Japan, to the APAN meeting in Nantou, Taiwan and to the Internet2 workshop in Denver, USA. For 3D transmissions, we used a portable ProjectionDesign projector with active glasses, which can use an ordinary projection screen and therefore does not require any arrangements from the meeting organizer. For 4K transmissions we used 4K TVs, which are now available from $500 (as of 2014).

It appears that Internet surgery streaming to individual doctors will be on rise, which can be accelerated by the increasing popularity of mobile devices, such as tablets and increasing mobile bandwidth, with the arrival of LTE (Long Term Evolution) services.

However, we believe that physical meetings at symposia bring additional value of direct information exchange among doctors. What turned very interesting for us with the technical background, is the stimulating role of medical symposia on technical requirements for surgery transmissions. The sheer fact that the doctors devote some of their time to come together results in the flood of new requirements of what else they want to see during the event to maximize the use of time. The first requirement is now to see multiple surgeries in parallel, performed at different institutions. This allows to ”skip” the less interesting parts and concentrate on the more interesting steps when they happen. The second requirement is to use multiple screens, usually one 3D screen switched to the surgery with the currently most interesting phase and several 2D screens for surgeon commentary, surgery room view, and other surgeries.

An example of a multi-screen event was the 19th Congress of the Slovak Society of Gynaecology and Obstetrics in Bratislava, Slovakia. We provided a real-time 3D transmission of the robot-assisted hysterectomy with bilateral adnexectomy and systematic pelvic lymphadenectomy performed by MUDr. Tibor Bielik, CSc. at the Faculty Hospital in Banska Bystrica. The surgery appearance in the lecture hall is shown in Fig. 4. The moderator was commenting on the currently most interesting phase of particular operations.

Fig. 4. Telesurgery from multiple hospitals to a congress in Bratislava

In Fig. 5 we show how real-time collaboration among countries can result in a cultural cyber performance. In the culture field, we have demonstrated interactions for remote control of models to access the national cultural heritage, we presented several chamber concerts in 3D vision or concerts with remotely playing musicians together with local performers.

This particular performance was named ”Dancing beyond Time” and involved approx. 100 people in three continents. The final event took place at the 36th APAN Meeting held in Daejeon, Korea on 21 Aug 2013. The event began at 08:55 UTC/GMT simultaneously in Salvador, Brazil (BR), Prague, Czech Republic (CZ), Barcelona, Spain (ES) and Daejeon, Korea (KR). The team included network engineers and researchers, audio-visual technicians, programmers, musicians, dancers, scene designers and choreographers, with some people spanning multiple areas. The music performance was captured by a 4K camera and delivered from HAMU to KAIST by a pair of FPGA-based 4K Gateway devices, which also provided a backward HD channel from KR to CZ for stage monitoring. Audio channels were transferred embedded in the video channel, which guaranteed a perfect video to audio sync in KR. The 4K video was sent uncompressed to preserve high quality.

Figure 5. Four-countries, three-continents real-time collaboration in musical and dance performance

The bitrate was approx. 5 Gb/s. The audience could thus observe a collaborative work of artists physically distributed in different continents.

V. CONCLUSION

The 4K Gateway was originally designed for 4K video contribution. Due to its very low added latency, it can be used for remote access to scientific visualizations, for medical sessions connecting operating theatres with lecture halls and conferences venues and for eCulture events and collaboration. It has been successfully used in various applications, which need high quality and low latency transmissions. In the future we plan to investigate the use of immersive visualizations for collaboration in performing arts, such as the CAVE devices (Cave Automatic Virtual Environment), involving other kinds of artistic expressions, such as fine arts and paintings and installations of more permanent infrastructures for the use in university lectures.

ACKNOWLEDGMENT

This work was supported by the CESNET Large Infrastructure project (LM201005) funded by the Ministry of Education, Youth and Sports of the Czech Republic and used multiple academic networks including Geant, TEIN, GLIF, Gloriad, KREONET, TWAREN and Internet2.

REFERENCES

[1] A. Ogawa, K. Kobayashi, K. Sugiura, O. Nakamura, J. Murai, *Design* and Implementation of DV based video over RTP, IEEE Packet Video Workshop, Cagliari, Italy, 2000.

[2] Tom DeFanti, Cees de Laat, Joe Mambretti, Kees Neggers, Bill St.Arnaud. TransLight: a global-scale LambdaGrid for e-science, Communications of the ACM, Vol. 46, No. 11, Nov. 2003, pp. 34-41.

[3] GLIF: Linking the world with light, Informational brochure, http://www.glif.is/publications/info/brochure.pdf.

[4] S. Shimizu, K. Okamura, N. Nakashima, Y. Kitamura, N. Torata, Y. Antoku, T. Yamashita, T. Yamanokuchi, S. Kuwahara and M. Tanaka, High-Quality Telemedicine Using Digital Video Transport System over Global Research and Education Network, Advances in Telemedicine:

[5] J. Halak, S. Ubik. “MTPP - Modular Traffic Processing Platform”, *Proceedings* of the IEEE Symposium on Design and Diagnostics of Electronic *Circuits and Systems*, 2009, Liberec, pp. 170-173.

[6] J. Navratil, M. Sarek, S. Ubik, J. Halak, P. Zejdl, P. Peciva, J. Schraml, Real-time stereoscopic streaming of robotic surgeries, Healthcom 2011, Columbia, Missouri, USA.

[7] J. Halak, M. Krsek, S. Ubik, P. Zejdl, F. Nevrela, Real-time long-distance transfer of uncompressed 4K video for remote collaboration, FGCS 27(7): 886-892, 2011.