I. Want. Pixels. (Entering the age of 4k)

Patrick Seeling and Martin Reisslein

Video has accounted for a large portion of the Internet network traffic in the past and is widely believed to continue to dominate in the foreseeable future. The Cisco Visual Networking Index by Cisco, Inc., for example, predicts that by 2017 more than two thirds of all consumer Internet traffic (including mobile) will be video. This astonishing amount of traffic is driven by the demand for more and more high-resolution media, on one hand, and the continuous improvements of network delivery capabilities allowing for streaming (e.g., a high-resolution video to a smartphone over cellular connections or an ultra–high-resolution video to increasingly affordable connected "smart" televisions). In addition to the capacity increases on the network side, several video coder/decoders (codecs) have been developed that allow for the efficient compression of video for network delivery or storage on high–capacity optical media. In recent years, we have witnessed a continuous increase in the resolutions that are supported by video codecs and requested by users. Figure 1 illustrates the size differences that are common today, from high–definition (HD) to the cinematic 4k resolutions of video.

Digital Object Identifier 10.1109/MPOT.2014.2335793
Date of publication: 13 November 2014

Simultaneously, a broad debate about the embedding of video playback capabilities into modern Web browsers has been waged with two main proposed venues of VP8/VP9 and H.264/H.265 video codecs. These proposals target the inclusion of video codecs into the HTML5 standards for video playback without the need for external plug-ins. The direct browser support will further increase HTML5 and HTTP video streaming; however, the actual long-term favorite codec to be universally supported across browsers has yet to pan out.

A brief comparison of video codecs

The VP9 video coding standard and software reference implementation were recently released by Google, Inc. and target high-resolution video. VP9 is likely to become extremely popular due to its license-free utilization in Web and mobile environments. In turn, this video codec has been integrated into many popular Web browsers or is in the process of being integrated, similar to the support of H.264/AVC, and is supposed to become the major codec in use by YouTube. The other current and widely used H.264/AVC and H.265/HEVC video codecs are also freely available in software reference encoder designs, which we utilize in our study as outlined in the next section.

Evaluation setup

We study the popular H.264/AVC and forthcoming (into the consumer space) H.265/HEVC and VP9 video codecs using the Creative Commons-licensed *Tears of Steel* short movie. *Tears of Steel* is one of the Blender open source movie projects, such as the popular

Big Buck Bunny movie, and depicts a futuristic battle between humans and robots. The movie has a wide range of content dynamics and is a combination of real video footage and overlaid rendered visual effects, which makes it representative for a broad range of recently generated commercial movies. We obtained the original 4,096 × 1,714, 16-b standard red, green, blue (RGB) color space image files of the individual video frames in the tagged image file format (TIFF) and converted them using `ffmpeg` into a 4,096 × 1,744 (4k)-resolution YUV color space sequence at 24 frames/s and a down-sampled 1,920 × 816 FHD version.

FIG1 An illustration of today's common video resolutions using a *Tears of Steel* screen capture: HD (720p, 1,280 × 720 pixels), full HD [(FHD) 1080p, 1,920 × 1,080 pixels], and 4k (2160p, 4,096 × 2,160 pixels).

We compare the encoding of the video sequence in both FHD and 4k resolutions with the different video encoders for simple, predictive coding (i.e., we do not consider B frames) and let the group of pictures' length for all video codecs equal 1 s (24 frames). In other words, we presume that every second an intracoded frame is sent, whereas the remaining video frames are all predictively coded from the directly preceding frame. We utilize a common video encoder configuration for the H.264/AVC and H.265/HEVC software reference encoders and the "best" single pass setting for VP9, providing an overall match of settings. The details of the encoder settings as well as the resulting video encoding data are available in video traces from the video trace library at http://trace.eas.asu.edu. The publicly available video traces enable statistical data analysis and video streaming simulations with the video encodings presented in this article.

Rate-distortion performance

First, we consider video delivery in native 4k resolution to a 4k TV set from the client–side point of view. The video quality is commonly measured as peak signal-to-noise ratio (PSNR), which in our case provides a logarithmic metric of the differences between the original (unencoded) video and the encoded video as a function of the average bandwidth requirement. We illustrate the resulting rate-distortion (RD) performance comparison in Figure 2 for bandwidths up to 10 Mb/s. We initially note that the 4k encodings in the newer VP9 and H.265/HEVC codecs are fairly close in terms of their performance; VP9 outperforms H.265/HEVC by approximately 0.5 dB PSNR or 0.5 Mb/s, depending on the point of view. The 4k encoding in H.264/AVC is significantly underperforming, with 3.5-dB loss in quality or severe additional demands in bandwidth to compensate for the lower coding efficiency of H.264/AVC. Clearly, this showcases the significant strides that were made in video codec development in recent years, now enabling the delivery of 4k video in good quality around 40 dB with around 3 Mb/s of bandwidth.

Current networks may need to deliver the 4k video stream at a smaller resolution, e.g., the FHD resolution, and then employ up-sampling by the TV set to a 4k display resolution. But how do lower resolutions fare when up-sampling is desired? Examining the up–sampled FHD video encoding results in Fig. 2, we note that with comparable bandwidths, the PSNR values indicate that it is always advisable to utilize the 4k full resolution. An interesting cut off range exists for the *Tears of Steel* sequence around 1–2 Mb/s; this is the highest compression point for the evaluated full resolution 4k encodings. In this range, it may be advisable to switch from a highly compressed 4k encoding to an up-sampled FHD version of the video sequence, as the highest 4k compression will likely yield an increased amount of visual artifacts. Joining both curves will hence give future streaming video service providers a full range of video delivery options, depending on client bandwidth availability.

Preliminary evaluations additionally suggest that there are significant differences in terms of computational complexity and resulting compression speed, with VP9 outperforming H.265/HEVC. Computational evaluations, however, have to be viewed with caution, as the reference software implementations are undergoing continuous development are currently not optimized for computational speed.

The Cisco Visual Networking Index predicts that by 2017 more than two thirds of all consumer Internet traffic (including mobile) will be video.

Future network bandwidth requirements

We now switch and consider the video service provider's point of view, which is to maximize the number of supported video streams with as little bandwidth as possible. We begin by assuming a very small data loss probability of 10^{-5}, which is commonly assumed to be low enough to allow cover-up (concealment) of losses by the client-side player. Next, we employ bufferless statistical multiplexing to determine

While the technological advances required for ubiquitous 4k video streaming are still being developed, the next technological revolution in the video streaming area is already emerging: three-dimensional (3-D) video.

the bandwidth needed to simultaneously stream 16 video sequences. We illustrate the relationship between the resulting minimum bandwidth requirement C_{min} and the PSNR video quality in Fig. 3. We note the reverse trend of the RD performance of the difference video codecs. While H.264/AVC encodings require the most bandwidth, H.265/HEVC and VP9 require significantly less bandwidth. Additionally, we observe that significant multiplexing gains exist, deducible when comparing the RD relationship in Fig. 2 and the resulting C_{min} in Fig. 3. For example, 16 of the lowest quality H.265/HEVC 4k encodings, each with approximately 1.2 Mb/s average bandwidth, can be statistically multiplexed using only around 6 Mb/s, which accounts for bandwidth savings of about 69%, and slightly lower 65% bandwidth savings for the lowest quality VP9 encoding in 4k resolution.

Following the PSNR video quality metric, we can derive an overall envelope as a rough guideline for streaming-video service providers as follows. For the low end of video quality/bandwidth requirements, it is advisable to use encodings in FHD (which also allows serving the existing FHD customer base as an added benefit) with H.265/HEVC as encoder, yielding highest compression gains. This is followed by the intermediate quality region, where a switch to 4k resolution with H.265/HEVC provides the best results (however for the existing customer base, continuation of the up-sampled FHD may make business sense, even at slightly reduced bandwidth savings). Lastly, for higher quality ranges, providers can switch to the VP9 encoder, which would yield the lowest bandwidth requirements for those quality levels. Additional considerations, such as adoption rates of codecs and their availability in cost–sensitive hardware implementations, may make a single solution more appealing for providers.

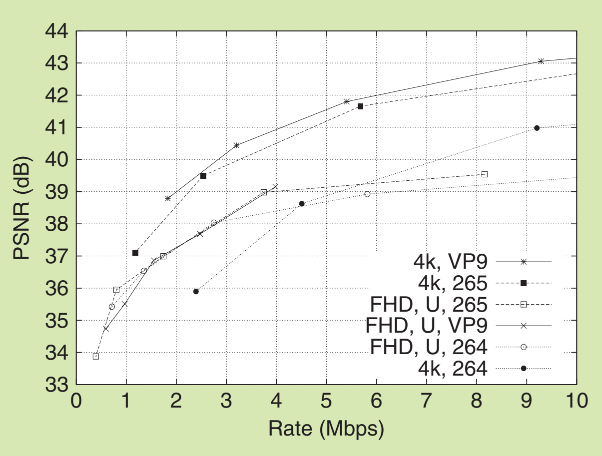

FIG2 RD performance of VP9, H.265/HEVC, and H.264/AVC codecs for native 4k and up-sampled FHD (up-sampled to 4k) versions of *Tears of Steel*.

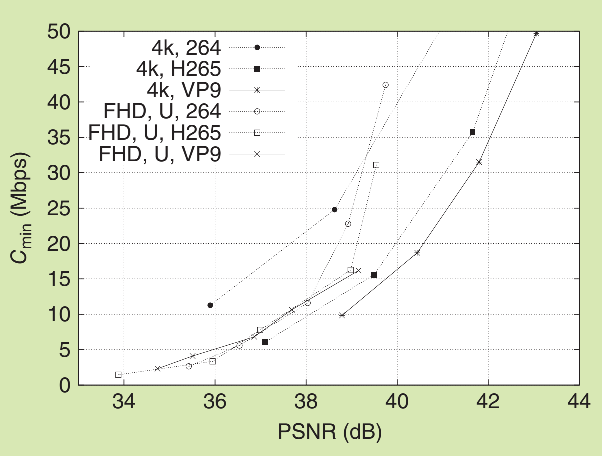

FIG3 Distortion-bandwidth requirement performance of VP9, H.265/HEVC, and H.264/AVC codecs for streaming 16 native 4k and up-sampled FHD (up-sampled to 4k) versions of *Tears of Steel*.

Conclusion

H.265/HEVC and VP9 video codecs exhibit comparable RD and statistical multiplexing performance at the 4k target resolution for the examined *Tears of Steel* video, which features a wide range of video content characteristics. Both H.265/HEVC and VP9 significantly outperform the preceding H.264/AVC codec. A high–level plan for streaming video service providers based on our observations is to send FHD versions to be up-sampled by the client (providing a basic quality) for low bandwidth regions and to switch to full 4k resolution encodings above the highest compression levels for that resolution. Some of the currently available 4k resolution TVs support both H.264/AVC and VP8 as predecessors of the recently finalized new standards H.265/HEVC and VP9 that we evaluated here. In turn, it appears likely that hardware vendors will support both H.265/HEVC and VP9 in the future to allow for broadest applicability of their products.

While the technological advances required for ubiquitous 4k video streaming are still being developed, the next technological revolution in the video streaming area is already emerging: three-dimensional (3-D) video. While the high data volumes arising from the step to 4k video resolution have video codec developers and video streaming providers scrambling, the next step to 3-D video will pose new challenges due to the need to stream multiple views of a video scene, each view in high (possibly 4k) resolution. Streaming a left view and a right view of a video scene could double the video bit rates. New emerging video codecs and 3-D video representation formats are trying to stem this renewed tide of rising video traffic. Stay tuned.

Read more about it

For overviews of the H.264/AVC, H.265/HEVC, and VP9 video codecs, we recommend:

- T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.
- G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, "The latest open-source video codec VP9: An overview and preliminary results," in *Proc. Picture Coding Symp.*, San Jose, CA, Dec. 2013, pp. 390–393.
- P. Seeling and M. Reisslein, "Video traffic characteristics of modern encoding standards: H. 264/AVC with SVC and MVC extensions and H.265/HEVC," *Scientific World J.*, vol. 2014, Article 189481, pp. 1–16, 2014.

For an introduction to video traces as descriptors for video traffic, we recommend:

- P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142–1165, 2012.

An introduction to statistical multiplexing of encoded video streams is provided in:

- G. van der Auwera and M. Reisslein, "Implications of smoothing on statistical multiplexing of H.264/AVC and SVC video streams," *IEEE Trans. Broadcast.*, vol. 55, no. 3, pp. 541–558, Sept. 2009.

For an introduction to the emerging area of 3D video streaming, we recommend:

- C. G. Gurler, B. Gorkemli, G. Saygili, and A. M. Tekalp, "Flexible transport of 3-D video over networks," *Proc. IEEE*, vol. 99, no. 4, pp. 694–707, Apr. 2011.
- A. Pulipaka, P. Seeling, M. Reisslein, and L. J. Karam, "Traffic and statistical multiplexing characterization of 3-D video representation formats," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 382–389, June 2013.

About the authors

Patrick Seeling (pseeling@ieee.org) is an assistant professor in the Department of Computer Science at Central Michigan University. He is a Senior Member of the IEEE.

Martin Reisslein (reisslein@asu.edu) is a professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University. He is an IEEE Fellow.

Digital Object Identifier 10.1109/MPOT.2014.2365052