

Response to Review Comments

Desktop and Mobile Web Page Comparison: Characteristics, Trends, and Implications

Original Manuscript Number: COMMAG-13-00421

T. Johnson and P. Seeling

January 2014

We would like to initially thank the editor and the reviewers for their very timely and in-depth review of our original manuscript. In the revision that accompanies this response document, we have carefully considered the suggestions and comments provided. For convenience, we have included the comments here in italics followed by our response.

Guest Editor

Comments to the Author:

Dear Author, the paper is really an interesting one, but as you can also see from the comments of the three reviewers it is not really ready to be published in IEEE Communication Magazine for which we require first class papers. However if you read carefully read and consider all the great feedback that the reviewers have provided I do believe it will become a great paper for the readers.

We are grateful for the opportunity and encouragement to improve this manuscript based on the thorough review comments. The new manuscript includes a section on error resilience, discussing the mechanisms mitigating the impact of packet losses. Given the tight space constraints for IEEE Communications Magazine manuscripts, it was not possible to further elaborate on error resilience, e.g., through specific quantitative experiments. Such a more detailed treatment of packet loss and errors would require its own manuscript dedicated specifically to this topic.

We have extensively expanded on the network transport aspects of the encoded video. Indeed, the traffic variability at the encoder output is less relevant in the wider context of network transport to buffered clients. We have explained this wider context of video network transport and illustrated the link bit rate requirements for the new video coding standards.

Reviewer 1

Comments to the Author:

typo: s/lesser extend/lesser extent/

We are grateful to the reviewer for noting the shortcomings of the original manuscript in considering only the RD characteristics and the VD characteristics at the encoder output. We have created a new Section III to specifically focus on the network transport of non-scalable video. This new section considers not only the RD and VD at the encoder output, but also the VD characteristics after a buffered smoother and the link bit rate requirements for a set of multiplexed video streams.

We have added a new subsection, namely Section II.D., to give a brief overview of error resilience strate-

gies. We do not specifically examine the detailed quantitative aspects of error concealment strategies, since such quantitative investigations would require significant space and could not be accommodated jointly with the basic video coding and network transport overview while meeting the tight space constraints of IEEE Communications Magazine manuscripts.

- The first paragraph of the introduction needs a reference. The authors describe a series of predictions regarding the increase of the mobile network traffic by 2014. This needs to be referenced.

These predictions are based on the Cisco Visual Networking Index. We have added the URL for this resource in a footnote, as it is not a formal publication. We would be glad to follow the editorial guidance as to how to best present this URL, e.g., whether the presentation in a footnote is acceptable or a different presentation, e.g., in parentheses in the text, would be preferable.

- The fundamental metric used in this paper to evaluate the different standards (VD), it is not referenced. Please reference the paper: P. Seeling and M. Reisslein, "The rate variability-distortion (vd) curve of encoded video and its impact on statistical multiplexing," IEEE Transactions on Broadcasting, vol. 51, no. 4, pp. 473-492, Dec. 2005. Or the corresponding seminal paper for this metric as it is not a conventional metric for evaluations as opposed to $R(D)$.

We agree that this Transactions on Broadcasting article presented the first formal definition of the rate variability-distortion (VD) curve. However, as the IEEE Communications Magazine as a tight limit on the number of references, we have added an alternate reference, namely Van der Auwera and Reisslein, "Implications of smoothing on statistical multiplexing of H.264/AVC and SVC video streams," *IEEE Trans. on Broadcasting*, 2009. This reference can cover the VD curve definition as well as the definitions of the delays introduced when smoothing video encoded with hierarchical B frames.

- There is no mention of packet loss. In a video communications network paper, some analysis regarding the robustness of the different standards to packet loss are necessary to provide the reader with a more complete vision of the state of the art. There should be a small section in the paper describing error concealment strategies (since they affect $R(D)$ values) that decoders use to conceal lost of information.

We agree that error concealment is an important aspect of video coding for communications. As recommended, we have added a subsection to the Overview of Video Encoding section that describes error concealment strategies, see Section II.D Error Concealment in the new manuscript.

- What is the reason why reference [15] (authors of this paper are part of [15]) uses 4 video sequences for the experimentation, and this paper only 2 for the 3-D section?

The main reason for considering only two video sequences was to fit curves for the four considered combinations of encoding (without or with cascaded quantization parameters) and streaming (aggregated or sequentially) into one plot for all presented video sequences. Following the recommendations of reviewer 3, we have removed the section on 3D video from the manuscript. Removal of this section has also helped

greatly in accommodating the greater level of detail requested in the reviews on non-scalable and scalable 2D video streams within the word limit of the IEEE Communications Magazine.

- Timing constraints due to dependencies in the decoder should have been addressed with greater extend (only end of section II mentions this). A figure/table showing the evolution of delays introduced by the different standards would be very helpful to assess their efficiency besides VD metric.

We have elaborated Subsection III.E. on Time Constraints due to Frame Dependencies to address the timing constraints with a greater extend. We have also added a table comparing the delays with classical B frame prediction and with hierarchical B frame prediction.

- A figure of a typical (block-based hybrid) video encoder/decoder block diagram should be included for visual interpretation as part of the introduction.

We are grateful to the reviewer for this suggestion. Indeed a block diagram figure helps in orienting the reader. We have added a block diagram of the entire video encoding, network transport, and video decoding system, see Fig. 1 in the new manuscript. Since this figure illustrates the entire video encoding/decoding and transmission system and accompanies a relatively brief overview of the video coding standards, it is less detailed in representing the encoder and decoder than block diagrams specifically focused on video encoding/decoding. Nevertheless, we believe that this added figure will help to orient the reader throughout the manuscript.

- At the end of each section the same conclusion is obtained: Newer standards offer better $R(D)$ performance but a high traffic variability cost, therefore strategies to mitigate traffic variability are needed. These conclusions can all be combined together in the last section saving some space that can be used to address some of the comments made above.

Correct, streamlining the discussions on conclusions saves space that is needed to address the preceding comments from this reviewer as well as the comments of the other two reviewers. We have entirely rewritten paper to achieve this streamlining.

Reviewer 2

Comments to the Author:

The topic of the paper, evaluating the trends of desktop and mobile web pages, is interesting. Especially the mobile dimension should make this study different from prior longitudinal studies. (I have not checked very carefully the prior studies so I don't really know how much they speak about the mobile aspects). However, there are several weak points in the present manuscript:

1. The paper claims that it is more up-to-date than prior studies. While that is probably true by looking at the dates of the sample it would have been important to compare how the results of this study differ from the prior studies e.g. from [4] or [5]. Such a comparison would have provided evidence to the authors statement that things are now different.

2. The major problem I find with the study methodology is that by blindly using the data in httparchive.org the authors draw conclusions which are probably wrong. As the authors observe the httparchive.org dataset had two major changes during the study period as the criteria for page selection were changed. Therefore I don't think it is fair to calculate the average growth by just taking the difference between the measurements in the first and last study dates and conclude that it shows the average growth rate. I suggest that the authors would (in addition?) take a smaller subset of pages that have been available since the beginning of the longitudinal study and see if they can see similar trends in that data. The two big jumps in the data can have a major corrupting effect.

3. Actually by looking at httparchive.org it seems that the authors have not done anything really radical, just analyzed the mobile data available on the site separately. While science does not need to be complicated having access to such good dataset would perhaps have allowed more thorough analysis and firmer conclusions.

4. While energy spending is definitely important for mobile pages I feel the authors emphasize it too much in the intro as their study does not deal with energy at all. Simply focusing on trends as they do in their measurements should be enough for a good and interesting paper.

5. By looking at httparchive.org it seems that the mobile data corresponds to iPhone pages and desktop to IE. It would have been good to state that (as other browsers may render the pages in a different way)

6. The discussion of Theil index is hard to follow. I first understood that with Theil index they compare how similar the desktop and mobile versions of the sites are, but by looking at Fig 2 with separate curves for mobile and desktop sites this cannot be the case. I am lost here and don't understand what this comparison is all about.

7. The data has some strange sections. E.g. in Fig 3a there are two major drops (in the middle of fixed curve) and a bump close to end. Discussing what these are would be interesting. Same in 3b close to end. Are they just random measurement errors or what?

8. On the last line of page 9 I don't understand what the authors mean by saying that increased use of frameworks has an influence on bytes per image

9. Fig 7 is a bit vaguely described. How are the curves for handsets, tablets, smartphones and North America estimated? Moreover, my personal experience over the study period is that popular mobile websites (in Finland) have become faster not slower.

This article attempts to provide a basic review of video coding standards and its transport for the communication and networking audience. However, the main conclusions of the article are misleading and experimental results are based on a limited set of test conditions. The terminology is also somewhat confused in places and the authors do not properly distinguish encoder choices and decisions from the design of the standard. For these reasons, the article should not be published. These issues are elaborated on further below.

We are grateful to the reviewers detailed insights and feedback on the original version of the manuscript. We have carefully considered the review comments in the development of the new manuscript, as detailed in the following.

The results reported in Figure 1 are based on CIF resolution video sequences. Such low resolution is no longer relevant; Internet and mobile streaming have moved well beyond this resolution. CIF is no longer even tested in the development of new standards. It is noted that the performance gap between HEVC and AVC is somewhat smaller than expected and not entirely consistent with the other expert reports, e.g., as given in [2].

We have removed all CIF resolution results from the manuscript. We have conducted new evaluations for video sequences in the full HD (1920 x 1180 pixel) resolution and report result from these full HD evaluations in the new manuscript.

For the full HD resolution, the performance gap between HEVC and AVC is indeed wider, as expected, consistent with the results reported in [2].

More importantly, it is asserted from the VD curves that advanced video coding standards have increasing traffic variability. This is one of the main conclusions of the paper and is a misleading statement. For one, it is not exactly clear how the results in Figure 1b were generated (e.g., prediction structures and QPs used), but clearly there are encoding strategies that could substantially reduce the traffic variability with minor impact on compression efficiency. For instance, the prediction structure could be changed to reduce variability and rate control could be employed as well to minimize bit rate fluctuations and satisfy any specified buffer constraints. It is quite surprising that such techniques are not even mentioned. Also, the authors only examine rate variability, but quality fluctuations are also quite important for certain applications and not at all reflected in the results provided or in the discussion.

We are grateful to these pertinent observations about the original manuscript. The original manuscript tried to cover too much ground, reaching from non-scalable video coding to scalable video coding (all for conventional 2D video) to 3D video encoding and streaming. For this new manuscript, we have narrowed the scope to non-scalable and scalable encoding and streaming of conventional 2D video. This allows to go into more detail on these topics within the tight space limits for the IEEE Communications Magazine.

In the new manuscript, we have entirely restructured the original section on non-scalable video coding and streaming to give more background on the different video transmission scenarios (as also suggested by Reviewer 3). The new Section II focuses on the overview of non-scalable video coding, while Section III focuses on the network transport of non-scalable video.

Within the scope of video transmission scenarios, we focus then on the case of video streaming and illustrate how smoothing and statistical multiplexing allows for effective transmission of video encoded using the latest coding advances, and how these coding advances translate into reductions of required network link bit rate (for a given PSNR video quality).

In general, the article also does a poor job in distinguishing encoder choices and decisions from aspects of the codec design. Figure 2 and the related text is one example of this. While hierarchical prediction structures are supported in AVC and HEVC, they need not be used. AVC and HEVC can also use the

classic IBBP prediction structures and have a very wide range of possible prediction structures including low delay prediction structures that also make use of B-pictures in an efficient way. Such remarks are repeated throughout the text. A non-expert reader could easily be led to believe that the prediction structure is something that is fixed and required to be used after reading this article.

Correct, the original article in trying to be brief in the discussion of the video coding background missed to make this distinction between encoder choices (for different video transmission scenarios) and aspects of codec design. We have added Section III.A. to make this distinction clear and highlight the flexibility and wide scope of covered video transmission scenarios by the latest standard codecs.

The authors also assert that the “higher traffic variability of modern coding standards can result in fewer supported streams compared to an older standard that has slightly higher mean bit rate but significantly lower traffic variability”. This seems to be quite an exaggeration, and again, misleading. Rate control will surely incur some loss in coding efficiency, yielding a bit stream with minimal rate fluctuations, but this loss is typically in the range of 5-10%. The compression efficiency gain of a newer generation standard is much higher than this. So, it is difficult to see how the argument put forward by the authors would ever hold up.

We agree that this statement in the original manuscript was based on a rather extreme rare case and was misleading for an overview article for the IEEE Communications Magazine readership. In the new manuscript, in Section III, we give a broad overview of the range of video transmission scenarios and then give illustrative results for the common case of video streaming with some buffering (smoothing) and statistical multiplexing, that is reflective of many practical video streaming settings.

The terminology that is used to refer to the various standards and extensions is also not accurate in several places. Again, this could create confusion for non-expert readers. The worst offense is probably the constant referencing of SVC (which is the scalable extension of AVC) in the non-scalable section! In the introduction, SVC is also used to denote the AVC standard for some reason and citations are made to [3],[4] rather than [13] or the more commonly cited T-CSVT paper from 2003 by Wiegand, et al. Also, the formal references to the standards by ISO/IEC and ITU-T are different than what is given in the paper. For instance, what people casually refer to as MPEG-4 Part 2 is formally ISO/IEC 14496-2.

We are grateful to the reviewer for noting the confusing terminology in the original manuscript. We have re-structured the terminology to be consistent throughout and to explain to the non-expert readers both the formal and casual terminology for the various video coding standards. Regarding the referencing of SVC and AVC, we added a clarification at the end of Section III.A. to clarify that we consider H.264 SVC in single-layer mode, which gives essentially equivalent encoding to H.264/AVC, for the illustrative results in the non-scalable video coding section. We have done this for two primary reasons: a) to illustrate the flexibility of the encoding standards, i.e., the point about the distinction between codec design and choices in usage of the codecs, and b) to have a consistent comparison benchmark across Sections III and V.

Additional comments:

- The introduction should acknowledge that there have been many studies over the past 15-20 years on long-term video traffic over networks.

Correct, there has been extensive research on long-term video traffic over the past 15-20 years. We have added a text remark to acknowledge this research in Section III.B., end of third paragraph. In order to stay within the 15 reference limit for IEEE Communications Magazine manuscripts, we have not included a reference with the remark. If the editor grants an exception to the 15 reference limit, we will be glad to support this remark with references.

- There is ongoing work on scalable and multiview extensions of HEVC that the authors may or may not be aware of. A stereo extension is expected to be finalized by January 2014, and responses to a call for proposals on scalable extensions were evaluated last year.

Thank you for this pointer. We have added a note on these scalable and multiview extensions of HEVC in the conclusion section to give an outlook to future developments.

- In MVC, the left and right views are part of the same access unit. It is not clear if the authors are taking this into account in their calculation of bit rate variability.

Thank you for the clarification on the access units in MVC. We have removed MVC from this new manuscript, thus the bit rate variability of MVC is no longer within the scope of this manuscript.

Reviewer: 3

Comments to the Author

First, I appreciate characterization papers - I always learn something new!

I feel the abstract and introduction conflate web content and internet content in the mobile space - and that's a hot topic about how much these two things interact. The study clearly deals with web content on mobile, but what might be more interesting is the content actually being consumed on mobile and the paper doesn't really give any insight into that.

III.A offers "continued increase in the average number of web objects requested likely results in additional networking overheads (such as connections setups, or DNS resolutions) and could have significant negative power consumption impacts". I really get nervous about phrases including "likely" and "could" when you are making the underlying argument of the paper. Do these things matter, and can you quantify how much? Common techniques such as persistent connections, SPDY, and speculative DNS can impact this data more than is immediately obvious. What's the right metric here - power? latency? speedindex? bandwidth charges?

in III.C you say "overall trend for both request origins is rather linear and close" and I have to say I think that's the most interesting conclusion in the paper - I would love to see it highlighted better! Similarly you say that the "average amount of bytes per image is fairly close" which I also found to be a non-intuitive and interesting finding. (and one I suspect doesn't hold up when talking about mobile apps

instead of mobile web apps - but it would be a great study to read).

in III.D you talk about "items cannot be cached effectively [...] due to limits below a day" and I think you may be missing a lot of the utility of caching. Many scenarios see most of their cache hits within that time frame (i.e. same session) as different portions of a site are thoroughly explored. Consider the number of transactions involved in that vs the small number involved in checking back on a different day for updates - short term cachability can be a serious boon.

in IV the discussion of download time and wait time doesn't seem to consider the more interesting metrics the industry uses for these things beyond total download page time. Time to first byte, time to first render, time to screen readable, and webpagetest.org speedindex are far more interesting metrics that don't necessarily track linearly against download time. Again, what is it about download time that is that interesting? I'm also concerned that the calculations are basically based on bandwidth calculations when web page download time is dominated by RTT more than bandwidth most of the time - so these headline numbers aren't very reliable but they are likely to stick with the reader as the take away of the article.