# A Media-Aware Transmission Framework for 3D-HEVC Over LTE Networks

Basak Oztas[1], Mahsa T. Pourazad[2], Panos Nasiopoulos[1], Victor C. M. Leung[1]
[1] Electrical and Computer Engineering, University of British Columbia
[2] TELUS Communications Inc.
basako@ece.ubc.ca, pourazad@icics.ubc.ca, panos@ece.ubc.ca, vleung@ece.ubc.ca

## ABSTRACT
The demand for high quality mobile video is driving advancement on many new technologies, from 4K to HDR to 3D. Content producers have begun creating high quality 3D video as well as building infrastructure for live streaming of stereoscopic 3D. When viewing live 3D video on a mobile device transmitted over a wireless network (such as LTE), network congestion and interference can lead to packet loss. The 3D viewing experience can be severely impacted by lost data, so it is vital to have a rate adaptation method for graceful degradation. We propose a combined system composed of a packet prioritization method for transmission and an error concealment strategy for viewing 3D-HEVC encoded video transmitted over an LTE network. This system improves the quality of experience by prioritizing packets with higher quality impact based on the number of dependent packets, leading to a superior PSNR and reduced playback freezes. This media-aware transmission approach is evaluated using the Vienna LTE-A Simulator; results show that it outperforms media-agnostic transmission from light to heavy network load with stringent startup delays. The evaluation demonstrates the utility of media-aware 3D-HEVC transmission and provides motivation for future media-aware developments in multi-view plus depth transmission.

## CCS Concepts
• **Information systems ~ Multimedia streaming** • **Networks ~Mobile networks** • *Computing methodologies~3D imaging*

## Keywords
3D video streaming; LTE-A networks; media-aware transmission; 3DTV; 3DVC; 3D-HEVC; multi-view video; rate adaptation; video-plus-depth.

## 1. INTRODUCTION
Video transmission accounts for the majority of internet traffic [1]. The success of streaming services has been driving demand for high quality video experiences, especially within the mobile domain. Many services already support 4K and stereo 3D viewing, and with recent advances in mobile network capacity and 3D compression, streaming video providers will be able to support multiple viewpoint video for perspective-corrected 3D displays and virtual reality headsets, for both home and mobile users.


*Conference'10*, Month 1–2, 2010, City, State, Country.

DOI: http://dx.doi.org/10.1145/12345.67890

In this paper we focus on the transmission of live 3D video represented in multi-view plus depth (MVD) format and transmitted as a highly-compressed video stream (taking advantage of multiple redundancies across viewpoints). This type of video is typically viewed on stereoscopic systems that require glasses to enable depth perception, or on a new generation of autostereoscopic displays employing parallax barriers or lenticular lenses to emit different pictures depending on the position of the viewer's eyes and do not require glasses for viewing.

The introduction of state-of-the-practice Long Term Evolution (LTE) cellular networks has provided sufficient bandwidth to transmit media with an increased minimum requirement, such as 3D video. However, mobile video demand is variable and peak demand cannot always be met with the available radio resources. Adverse channel conditions, caused by factors such as fading and interference, lead to high packet loss or require a conservative Modulation and Coding Scheme (MCS); this in turn hinders viewers' Quality of Experience (QoE) due to freezing, rebuffering or lower picture quality. Continuous playing and consistent perceptual quality are important for a pleasant viewing experience [2]; as little as 1% of a video freezing can reduce video consumption by 5% and a startup delay longer than 2s can significantly increase user abandonment [3]. Given currently available technologies, it is difficult to guarantee a sufficiently high level of QoE for users viewing MVD video, since the expected load is substantially greater than 2D video (in our own simulation setup, two views with depth maps increase the bandwidth requirement by 60% compared to 2D video). In the

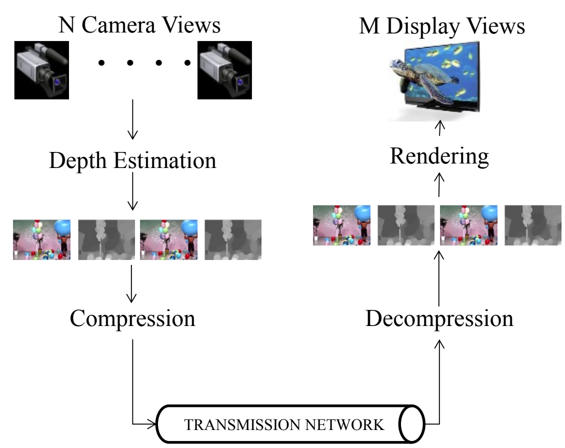


**Figure 1. Infrastructure of multi-view plus depth video from capture to viewing**

case of 3D-HEVC streaming, a high bitrate, low latency (as in any video streaming application) and tolerable packet loss are desired to avoid artifacts in 3D video viewing.

MVD video can be highly compressed for transmission using the 3D High Efficiency Video Coding (3D-HEVC) standard [4]. HEVC compresses video by creating dependencies which exploit temporal redundancies across frames [5]. The 3D-HEVC extension further exploits additional redundancies among different multi-view streams which capture the same scene from multiple viewpoints (typically highly similar video streams) to achieve more efficient compression, compared to simulcast streams. However, with an increase in compression comes an increasingly negative impact as a result of frame packet loss. This would be additionally jarring if the packet loss affected the perceived depth of the viewed scene, which is difficult to avoid since the dependencies are no longer just temporal or intra-picture.

To facilitate higher quality 3D-HEVC transmission we propose to extend the LTE network with an additional media-aware application layer transmission protocol. Previous methods using media-aware protocols have been used for 2D and simulcast 3D but are not directly applicable to 3D-HEVC encoded MVD due to newly introduced interlayer dependencies and the possible involvement of rendering at the receiver end. The contribution of this work is a method that, given any network condition, (especially those that limit available bandwidth per user) improves the QoE of the transmitted video by prioritization of video packets through analysis of the inter-packet compression dependencies and gracefully degrades the decoded video by concealing the loss of less important packets. Dependency analysis provides a detailed measure of importance of the video packets, which was not employed in the previous approaches in packet prioritization with 3D-HEVC [6]. This method is also convenient for network transmission and can be employed in other transmission technologies as it is performed entirely in the application layer and does not require modifications in the lower layers of the LTE, unlike previous methods [6].

Our novel algorithm is evaluated within the Vienna LTE-A Downlink System Level Simulator against a media-agnostic method, measuring the PSNR, and the frequency and duration of playback freezing. The transmission performance was tested in unreliable network deployment scenarios and evaluated in terms of QoE, which has not been performed before. We would like to achieve a persistently high QoE for end users of 3D video streaming applications under adverse network conditions. The main idea behind rate adaptation is that the video bitrate should be adapted on the fly to provide uninterrupted and smooth playback on the client side. This allows for graceful degradation of video quality while preventing the video playback from re-buffering and freezing when network resources are scarce. Our error concealment strategy in conjunction with the media-aware transmission ensures a higher QoE for adverse network conditions.

The following section summarizes the related work within LTE, 3D video and 3D-HEVC. The packet prioritization and error concealment methodology used to improve the user QoE is presented in Section 3. We then describe our evaluation to test the media-aware approach against a media-agnostic algorithm in Section 4. The paper is concluded with a discussion on the use of media-aware transmission of 3D packets in Section 5.

# 2. BACKGROUND

Today's playback technologies range from traditional monoscopic displays to multi-view autostereoscopic displays. For example, a traditional monoscopic display needs a single colour (texture) video, whereas two texture videos are needed for a stereoscopic display. For more diverse viewing experiences such as flexible stereoscopic baseline support for varying viewing conditions [7] and for autostereoscopic displays and free-viewpoint video, depth maps can be coupled with the texture video and new virtual views can be synthesized at requested positions with techniques such as DIBR [8]. Multi-view plus depth (MVD) is a video representation format that can satisfy the demands of such diverse playback technologies. To produce this format, a scene is captured with one or more cameras and depth information of the scene can be retrieved through depth cameras or generated with texture-based reconstruction [9], as shown in Figure 1. Each conventional texture video is paired with its associated monochromatic depth video to enable rendering of the virtual views at the receiver side. Typically, one to three viewpoints are sufficient to synthesize a wide range of virtual viewpoints.

Even with HEVC codecs, the simulcast streaming of MVD format consumes a large amount of bandwidth, and a specialized compression scheme was necessary to take advantage of the redundancies between the viewpoints. 3D-HEVC was created as an extension to HEVC compression standard [4] by joint efforts of MPEG and ITU VCEG. It achieves high compression rates by taking advantage of the similarities between views, and within the components (texture and depth) of the same view. The extension also adopts new low level tools that are specifically designed to exploit the characteristics of depth maps. The standard refers to each component of each view as a layer. The base layer (i.e. view 0's texture) must be compressed with a conventional HEVC encoder for backward compatibility. All layers belonging to the same instant are multiplexed to create an access unit (AU), a concept similar to its HEVC counterpart. A sample configuration in Figure 2 demonstrates the different types of dependencies between the layers. These new hierarchical dependencies between layers offer new challenges to the video transmission community.

There have been remarkable developments in the literature which underpin today's video streaming technology. Researchers have begun exploring the possibility of streaming MVD over LTE [10]. Even though rate adaptation is very well studied in the literature for 2D video [11], its scope has been mostly limited to the stereoscopic case for 3D content.

There have been a few studies investigating the factors affecting

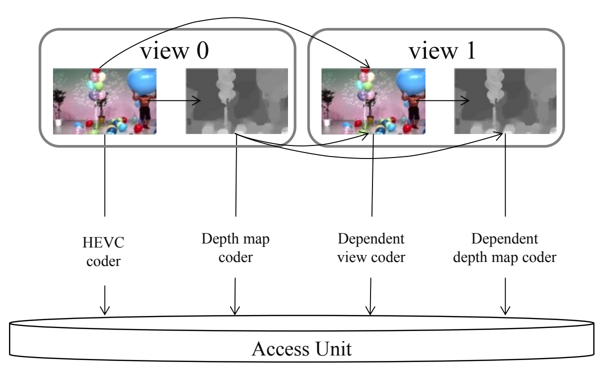


**Figure 2. A sample dependency hierarchy among layers of 3D-HEVC**

QoE of 3D video within the transmission context. One series of experiments was conducted assessing the effects of rate fluctuation on quality of experience (QoE) for frame compatible stereoscopic format [2]. It was reported that freezing and rebuffering are detrimental to QoE, whereas SNR based quality adaptation is a more preferable option. Another finding was that when the bit-rate dropped, viewers preferred switching from 3D to 2D. A similar study conducted subjective experiments to evaluate the effects of various adaptation strategies on simulcast stereoscopic video [12]. The evaluated strategies were SNR change, asymmetric coding, temporal resolution reduction, 2D⇔3D switching and various video freeze durations for stereoscopic content. Results show that a higher frame rate in the original video implies a higher resilience to frame rate drops. Their results also suggest that switching to 2D video can be more favourable in cases where the encoded picture quality is not sufficient to provide a pleasant 3D experience. Freezes were also shown to be detrimental to QoE, especially when the content is streamed in real time and cannot be resumed from the same playback point. The level of annoyance experienced by viewers is correlated with the length of the freeze duration. Although these tests are conducted on stereoscopic video, the effects of 3D⇔2D switching and video freezing/rebuffering provide valuable insights into QoE in MVD video. In another study, a novel rate adaptation concept specifically targeting MVD applications was put forth [13]. It is claimed that rate adaptation can be achieved by varying the stereoscopic viewing baseline (the distance between left and right camera positions). A small decrease in the 3D effect can successfully transform into a large increase in the overall user experience for small number of transmitted views [13].

In traditional video streaming, packets are transmitted according to their decoding order, which is known as earliest first scheduling. Instead of this default behavior, employing a server-side packet scheduler to achieve optimal rate distortion is shown to be a potential approach to rate adaptation [14-16]. A priority-driven packet scheduling technique for SNR layered video has previously been proposed [14], where packets are ordered based on their contribution to the overall video quality. Essentially the base layer of the encoded video is guaranteed to have an acceptable frame rate before the transmission of enhancement layers. This method bypasses the bandwidth estimation process and achieves rate adaptation effortlessly through temporal and SNR scaling. In an alternative method, layers of scalable video are retrieved from the server in the order of their contribution to the overall quality [15]. This helps maintain the best achievable quality under adverse network conditions and smoothes out the quality over a given interval. An apparent challenge pertaining to priority buffering is the selection of a buffer filling strategy. This problem was previously tackled for SVC coded video [16], however, further complications emerge in the case of MVD video due to the convoluted nature of inter-view dependencies and the rendering process.

# 3. PROPOSED APPROACH FOR MEDIA-AWARE TRANSMISSION

The state-of-the-practice for transmitting 3D-HEVC content over LTE networks is to treat all packets equally within the stream, for all connected users. A media server encodes (or uses pre-encoded) MVD video in 3D-HEVC and provides it to the eNodeB, which will serve a user with the encoded video. The media server does not adapt its rate according to current network conditions and has no knowledge of the current throughput of the user devices. For a media-blind eNodeB, the packet scheduler attempts to send the packets in the order in which they arrive from the media server (generally the encoding order). This ordering is also known as earliest first scheduling. If adverse network conditions occur and the available bandwidth becomes inadequate for various reasons such as heavy traffic load due to a large number of users or high interference, then network congestion will cause packets to be dropped at the eNodeB: this can occur due to a packet missing its playout deadline or transmission failure. Dropping packets without considering the characteristics of the compressed media stream can lead to an inefficient use of available bandwidth. Additionally, the received stream will contain artifacts after the packet loss, causing 3D video glitches and re-buffering of the video. When operating under a strict playout delay, the user may experience many long pauses in the streamed video.

Our proposed algorithm for increasing the quality of viewed 3D-HEVC content transmitted over LTE networks consists of two parts. First, we order the upcoming packets based on the number of dependent packets they have in the packet buffer of the scheduler, and prioritize highly-referenced packets whenever bandwidth is available. Sending packets which are referenced more will increase the overall viewing experience at the expense of losing less-important pictures. Second, we employ an error concealment strategy to increase the QoE for the user by providing graceful degradation in the presence of packet loss. This has limited potential in the case of arbitrary packet loss, but is designed to take advantage of the prioritized packets to decrease the perception of interruption. The target conditions of this algorithm are for viewing live streams with a stringent playout delay within a congested network.

Previous work promoted the Modulation and Coding Scheme (MCS) for I-frames to one with a lower packet loss probability [6], however this does not take into account non I-frame packets which have a high number of dependencies. Moreover, it can deteriorate an already congested network by decreasing the efficiency of bandwidth usage because of the adoption of a more reliable MCS. Our proposed framework provides an opportunity for packets to be individually prioritized based on their utility for decoding (even if their own playout deadline has passed), without enforcing cross-layer changes on the LTE system.

## 3.1 Content Packetization

The input videos are encoded using 3D-HEVC with all texture and depth layers, using one slice per picture per layer. Each set of pictures to be viewed at the same time (i.e. all layers with the same playout deadline) are contained within one access unit (AU) and have the same picture order count (POC). Dependencies are encoded within each AU as in Figure 2, and the texture of view 0 is designated as the base layer (layer 0). Each layer within an AU is packetized individually. Each packet is given a *transmission deadline* corresponding to the playout time of its dependent packet with the highest POC; even if a packet's playout delay passes, the scheduler may still attempt to send it if it is useful to later packets such as in the case of an I-frame.

## 3.2 Packet Prioritization

Our packet prioritization method is based on the hierarchical importance of each packet and sends more important packets within the buffer first. The packet importance metric is defined based on the number of dependents that packet has. For example, I-frames have many dependents and therefore tend to be the most important frames to send. The base layer within each AU is a dependency for every other layer in the AU, therefore is designated as more important than the other layers under our importance metric.

To reduce the complexity within the media-aware LTE system, we compile *dependency lists* during encoding. A single dependency list for packet $p_n$ is the set of all packets in the content which refer to (i.e. depend on) $p_n$ both directly and indirectly. Direct references are specified in the reference lists; indirect references include inter-layer dependencies for 3D-HEVC texture-and-depth coding (such as the depth layer of one view depending on both the texture and the depth of another view) and recursive dependencies.

However, using all dependencies of a packet as a measure of importance can lead to overcompensation, i.e. abandoning packets which are unimportant yet urgent. This is due to highly important packets later in the stream entering the packet buffer and dominating the urgent packets. To mitigate this, we only consider dependencies as relevant for measuring importance if they are in the packet buffer. In this case, if an I-frame enters the packet buffer but none of its dependents are in the buffer yet, it will not be considered important. As more of its dependents enter the packet buffer, its importance will gradually increase.

## 3.3 Packet Failure

For 3D-HEVC content, packet failure can lead to a significant degradation in the quality of the viewed video due to the involvement of rendering process and highly complicated hierarchical encoding structure. Failure occurs in our framework due to network failure (NACK) or through the expiry of a packet after its playout deadline.

Dropped packets due to network failure occur most often in larger packets (since they are comprised of many transport blocks). These packets are depended upon by a large number of other packets for successful decoding, therefore their arrival is crucial. At the cost of using extra bandwidth, our system enforces reliability through retransmission of lost blocks to ensure video packets are received. Due to the extra time taken for re-transmission, there may be a slight increase in the number of expired packets. This ensures that a packet is marked as failed only when the transmission deadline expires.

## 3.4 Error Concealment Strategy

The second contribution of this work is a method for rendering decoded 3D-HEVC content with missing packets through error concealment. This strategy is applicable to any transmission method; however, it will provide a higher quality with our prioritization scheme as we demonstrate in the results. The error concealment method is presented for two views along with their corresponding depth maps, however it can be applied to a higher number of views by extending the recovery rules to cover the compressed number of views.

For each decoded AU (all pictures with the same POC), we apply one of the following strategies:

- Interpolated 3D: All views (textures and depths) within the AU were successfully decoded; View 0 is played as the left view and a synthesized right view is rendered via interpolation between View 0 and View 1. This provides the maximum possible quality since all packets were successfully delivered and decoded.
- Extrapolated 3D: View 0 was successfully decoded however View 1 could not be decoded, due to packet loss. View 0 is played as the left view and a synthesized right view is rendered via extrapolation of View 0.
- 2D: The texture of the base view (Layer 0 of the AU) was the only successfully decoded layer. The decoded texture picture is shown to both left and right eyes of the viewer, i.e. the viewing is switched to 2D. This mode is equivalent to watching monoscopic video.
- Frame Copying: None of the layers of the AU could be decoded. The left and right pictures from the last decoded AU (which could be any of the three modes above) are copied to the current frame. In QoE terms, this can be perceived as a drop in frame rate or freezing/rebuffering depending on the duration of this mode.

# 4. EVALUATION

The method detailed in the previous section is tested in two configurations within an LTE simulator: traditional media-agnostic transmission and our proposed media-aware packet prioritization transmission. The main goal of the novel media-aware method is to provide the optimal QoE under high network load or adverse network conditions. The configurations we use for the simulator target a variety of network conditions with a focus on congestion. We evaluate the results based on metrics of Quality of Experience (QoE).

Under a media-aware LTE network transmitting 3D-HEVC video we hypothesize the following: the QoE should improve via the media-aware packet scheduler that considers the impact of each packet on the viewing quality. Dropped packets should be of lesser importance (i.e. fewer references), therefore the video can be intelligently rebuilt using the error concealment strategy to avoid severe impact on the user experience. The QoE can be measured through analysis of the PSNR and by a reduction in the frequency and duration of rebuffering due to adaptive streaming.

## 4.1 Experimental Setup

The media-agnostic and media-aware transmission methods were tested on the ‘Balloons’ video [17] from the MPEG 3DTV Call for Proposals [18]. The 3D-HEVC encoded video was transmitted over a modified version of the Vienna LTE-A Downlink System Level Simulator [19], executed in parallel on a grid compute server. The details are provided in the following sections on video, LTE and simulation setup.

### *4.1.1 Video Setup*

The Balloons sequence was encoded using the 3D-HEVC encoder reference software HTM 14.0 [20], with two views and two depths using Cameras 1 and 5; sequence details are in Table 1 and the bitrate breakdown per layer is in Table 3. The video stream is packetized at the layer level, i.e. each packet contains precisely one layer from one AU.

The viewing setup for the final rendered video plays decoded Camera 1 on the left and synthesized Camera 3 (from available decoded views) on the right. When only the texture of Camera 1 is decoded successfully, our graceful degradation method uses decoded Camera 1 for both left and right (a switch to 2D viewing) to avoid re-buffering. VSRS 3.5 [21] is used to synthesize virtual views in Camera 3’s location. Camera 3 was chosen as the virtual view position so we could perform objective measures in a miss-one-out fashion i.e. evaluating PSNR between the synthesized Camera 3 and the original uncompressed version.

In this study, we focus on the dynamics between the hierarchical structure of a 3D-HEVC compressed video and we expect that the results will be sufficiently representative of a wide variety of content encoded with a similar structure.

The target application for this work is live streaming, therefore the delay can be larger than in applications such as conversational video. Based on a study showing that users abandon video with a

**Table 1. Video and compression parameters**

**Video Sequence Parameters**

| | |
|---|---|
| Frame size | 1024x768 |
| Frame rate | 30 fps |
| Video length | 10 sec |
| Encoded views | Base view: Camera #1<br>Dependent view: Camera #5<br>Both texture and depth |
| Camera Spacing | 5 cm |

**Compression Parameters**

| | |
|---|---|
| I frame (CRA) period | 32 (~ every 1 sec) |
| GOP size | 8 |
| GOP structure | Random access (IBB... IBB…) |
| Slicing | 1 per frame |
| Quantization parameter | Texture: 25 Depth: 34 |

startup delay of more than 2 seconds [3], we tested startup delays of 0.5s, 1s and 2s.

### *4.1.2 LTE Setup*

Our method was integrated into and evaluated using the Vienna LTE-A Downlink System Level Simulator v1.8 [19]. This simulator provides a detailed simulation of system level performance of an LTE network, where the physical layer is abstracted to allow the analysis to concentrate on the network performance. For our study, the input video bitstream is provided to the virtual LTE eNodeB which then sends the sequence to the users under varying network loads.

The LTE simulator is configured to model a bandwidth of 3MHz using a single antenna, proportional fair scheduling, and stationary motion for the users. The eNodeBs and sectors are arranged in a regular hexagonal grid structure, with macroscopic pathloss model TS36942 [22] using the urban setting. The number of users modelled in the simulations varies between 12 and 24 to represent moderate to heavy network load.

### *4.1.3 Simulation Setup*

The simulations were performed using the Vienna LTE-A Simulator in Matlab v8.3 and executed on the Compute Canada WestGrid HPC system. Processing times varied between ~0.5 hours per simulation to ~4 hours per simulation, depending on the

**Table 2. LTE simulator configuration**

| | |
|---|---|
| Transmission mode | Single antenna |
| Scheduler | Proportional fair scheduler |
| User mobility | Stationary |
| System bandwidth | 10 MHz |
| Carrier frequency | 2 GHz |
| Link-to-system interface | MIESM |

**Table 3. Bitrate breakdown per camera and layer**

| | Texture | Depth |
|---|---|---|
| Camera #1 | 845 kbps<br>43.2 dB | 129 kbps<br>40.6 dB |
| Camera #5 | 336 kbps<br>41.9 dB | 94 kbps<br>38.78 dB |
| | Total of four layers: | 1409.0 kbps |

number of users (a larger number of users led to a longer simulation). Eighteen different configuration combinations were tested, each performed ten times with different seeds to verify consistency. Table 4 shows the different conditions tested. For each condition, the start time for each user's session is chosen randomly within the duration of one I-frame period; this distributes the traffic more evenly and prevents I-frame alignment, modelling a real-world viewing of multiple independent streams. The simulated duration was equivalent to the length of the original video sequence (10s) plus the playout delay (0.5s-2s) plus the random start time (total duration varies between 10.5s and 13s).

## 4.2 Results

Based on the setup described above, we collected data examining the quality of experience of 3D-HEVC video live streamed over congested networks. Our novel media-aware method outperforms the traditional media-agnostic approach in all tested conditions. Under relatively light network load, both methods are able to provide high quality playback (based on PSNR) as shown in Figure 3. Under heavier network load, more packets are dropped which leads to a significantly lower PSNR for the agnostic approach. With media-aware transmission, packets with lower importance are lost and the error concealment methods can be employed to greater effect to achieve a higher PSNR. The improvement in PSNR can be seen in Figure 3 with a network load from 12 to 24 users with a startup delay of 0.5s, 1s and 2s.

The number of copied frames for error concealment (due to a failure to receive a decodable AU) was similar for both methods (as shown in Table 4), which mostly depends on the available network bandwidth. As can be seen in Figure 4, the mean duration of consecutive dropped AUs (frozen playback) for agnostic transmission is in the order of hundreds of milliseconds, which

**Table 4. Percentage of video frames with frame copy concealment (freezing/rebuffering) for all conditions. Agn. stands for media agnostic, and MA stands for the proposed media-aware technique**

| | Watch Delay (ms) | | | | | |
|---|---|---|---|---|---|---|
| | 500 | | 1000 | | 2000 | |
| # Users | Agn. | MA | Agn. | MA | Agn. | MA |
| 12 | 11% | 8% | 9% | 8% | 5% | 7% |
| 18 | 15% | 13% | 12% | 11% | 9% | 10% |
| 24 | 31% | 27% | 26% | 23% | 19% | 21% |

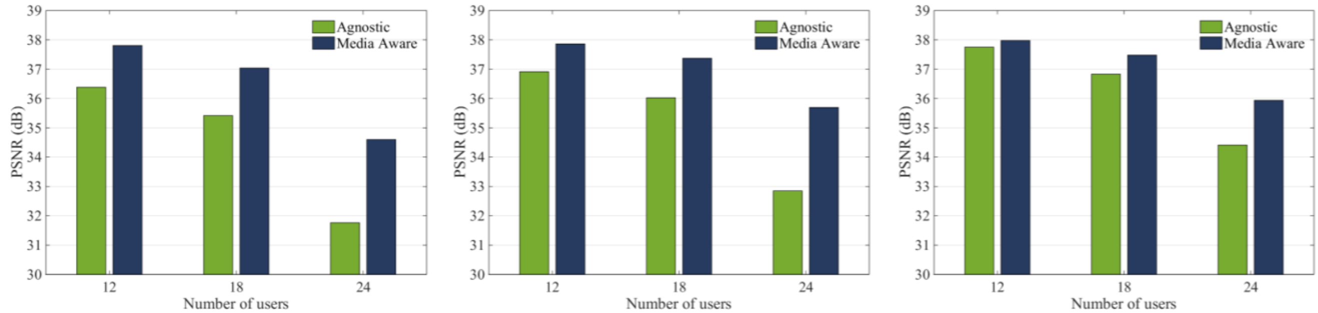


**Figure 3. Average PSNR value for varying number of users and startup delays of 500ms, 1000ms, and 2000ms (left to right)**

would interrupt the viewing experience and would be clearly visible to the viewer. For media-aware transmission, the mean consecutive duration is in the order of tens of milliseconds, which will appear as a lower frame rate to the users and will not be perceived as rebuffering. The media-aware method optimized the packet loss for higher layers and was able to recover more quickly; this led to a mean freeze duration for the media-aware method which was 15 times smaller than the equivalent for the traditional agnostic method, as illustrated in Figure 4. Previous work has shown that video playback with 1% freezing leads to 5% less viewing of the video [3], highlighting the importance of maintaining a continuous playback.

The mean consecutive freeze durations in Figure 4 for a startup delay of 0.5s appear to be much lower than for the higher delays. This value is calculated based on the duration of each freeze: for the case with the shorter startup delay, freezing occurs more frequently but with a shorter average duration. For the case with the longer delay, there are fewer, longer freezes. The mean duration of the freezes appears to be due to the size of the startup delay (also known as buffer length), since the packet buffer under congestion attempts to send earlier packets for a longer time, and after failing, has to wait longer for impactful packets.

From the different startup delay conditions presented in Figure 3, it can be seen that the performance converges as the startup delay increases. This is expected given that the larger buffer allows for improved recovery from delay in the network. The difference in performance is highlighted under the most challenging conditions, with a startup delay of 0.5s and the most congested network (24 users), where the media-aware transmission provides a final video quality 2.9dB PSNR higher than the agnostic transmission (34.6dB PSNR vs 31.7dB PSNR).

# 5. CONCLUSION

We have presented a media-aware transmission method which prioritizes packets based on their importance as well as an error concealment strategy, which attempts to reconstruct the lost parts of the stream. The proposed approach leads to an objectively higher quality of experience even with stringent startup delays and in the presence of adverse network conditions. Results show that our novel packet prioritization method leads to a significantly higher objective quality (PSNR) after transmission and decoding by dropping less important packets. The quality of experience is substantially improved due to the order of magnitude decrease in the mean duration of playback freezes: 15 times shorter for media-aware transmission than agnostic. Through our error concealment strategy this would be perceived as a lower frame rate instead of rebuffering, demonstrating a graceful degradation and achieving rate adaptation on the fly instead of having to pre-encode the same content at multiple bit-rates or having to estimate available throughput.

We have shown that our method provides for service to more users than with a media-agnostic approach. Based on the quality of video received using our method, as many as 24 users can receive watchable content with a 0.5s startup delay. The agnostic method fails to provide watchable video at this level, and increasing the network load may lead to a low quality of experience for users with the media-aware approach as well. Increasing the watch delay would likely allow for more users to be added to the network. However, a watch delay longer than a few I-periods may lead to over-prioritization of large frames that are nonurgent, e.g. a newly arriving I-frame may be sent before an urgent packet with relatively low importance. We intend to augment the method to use viewing deadlines as part of the optimization strategy and allow longer watch delays and shorter I-periods.

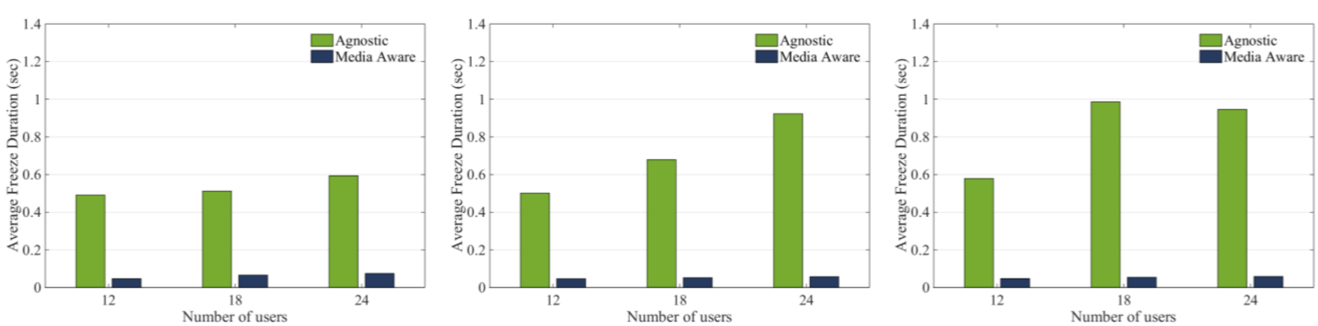


**Figure 4. Average freeze duration for varying number of users and startup delays of 500ms, 1000ms, and 2000ms (left to right)**

## 6. ACKNOWLEDGMENTS

This work was supported in part by the NSERC DIVA Strategic Research Network, TELUS and other industry partners.

## 7. REFERENCES

[1] Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update*. 2011-2016.

[2] Tavakoli, S., Gutierrez, J., and Garcia, N. 2013. Quality assessment of adaptive 3D video streaming. *Proc. SPIE Three-Dimensional Image Processing and Applications*, vol. 8650, 2013.

[3] Krishnan, S.S., and Sitaraman, R.K. 2013. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking*, 21(6), pp.2001-2014.

[4] Tech, G., Wegner, K., Chen, Y. and Yea, S. 2015. 3D-HEVC Draft Text 7, document JCT3V-K1001. Geneva, Switzerland, Feb.

[5] Pourazad, M. T., Doutre, C., Azimi, M., and Nasiopoulos, P. 2012. HEVC: The new gold standard for video compression: how does HEVC compare with H.264/AVC? *IEEE CE Magazine*, vol.1, no.3, pp.36-46, Jul 2012.

[6] Jassal, A., Oztas, B., Pourazad, M.T. and Nasiopoulos, P.. 2015. A packet prioritization scheme for 3D-HEVC content transmission over LTE networks. *In 2015 IEEE International Conference on Communication Workshop (ICCW)* (pp. 1788-1793).

[7] Shibata, T., Kim, J., Hoffman, D., and Banks, M. S. (2011) The zone of comfort: Predicting visual discomfort with stereo displays. J Vis July 21, 2011 11(8): 11

[8] Fehn, C. 2004. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. *Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI*, 93 (May 21, 2004)

[9] Kauff, P., Atzpadin, N., Fehn, C., Müller, M., Schreer, O., Smolic, A., and Tanger, R. 2007. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. *Image Commun.* 22, 2 (February 2007), 217-234.

[10] Piro, G., Ceglie, C., Striccoli, D., and Camarda, P. 2013. 3D Video transmissions over LTE: A performance evaluation. Proc. IEEE EUROCON, pp. 177–185, 2013.

[11] Nur, G., Arachchi, H.K., Dogan, S., and Kondoz, A.M. Kondoz. 2012. Advanced adaptation techniques for improved video perception. IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 2, pp. 225–240, 2012.

[12] Gutierrez, J., Perez, P., Jaureguizar, F., Cabrera, J., and Garcia, N. 2012. Subjective study of adaptive streaming strategies for 3DTV. Proc. IEEE International Conference on Image Processing, pp. 2265–2268, 2012.

[13] Oztas, B., Pourazad, M. T., Nasiopoulos, P., Sodagar, I., and Leung, V. C. M. 2010. A rate adaptation approach for streaming multiview plus depth content. *Computing, Networking and Communications (ICNC), 2014 International Conference on*, Honolulu, HI, 2014, pp. 1006-1010.

[14] Kuschnig, R., Kofler, I., and Hellwagner, H. 2010. An evaluation of TCP-based rate-control algorithms for adaptive Internet streaming of H. 264/SVC. *Proc. ACM SIGMM Conference on Multimedia Systems*, pp. 157–168, 2010.

[15] Schierl, T., Sanches de la Fuente, Y., Globisch, R., Hellge, C., and Wiegand, T. 2011. Priority-based media delivery using SVC with RTP and HTTP streaming. *Multimedia Tools and Applications,* vol. 55, no. 2, pp. 227–246, 2011.

[16] Andelin, T., Chetty, V., Harbaugh, D., Warnick, S., and Zappala, D. 2012. Quality selection for dynamic adaptive streaming over HTTP with scalable video coding. *Proc. ACM SIGMM Conference on Multimedia Systems,* pp. 149–154, 2012.

[17] Tanimoto, M., Fujii, T., Tehrani, M. P., and Wildeboer, M. 3DV/FTV EE1 report on Kendo and Balloons sequences. ISO/IEC JTC1/SC29/WG11 M17207, Nagoya University – Japan.

[18] *Call for Proposals on 3D Video Coding Technology.* ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, Geneva, Switzerland, March 2011.

[19] Ikuno, J. C., Wrulich, M., and Rupp, M. 2010. System Level Simulation of LTE Networks. *Vehicular Technology Conference (VTC 2010-Spring) 2010 IEEE 71st*. Taipei, 2010, pp. 1-5.

[20] 3DV HEVC Test Model (3DV-HTM) version 14.0. Retrieved: July 2016 [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-14.0/

[21] Tanimoto, M., Fujii, T., Suzuki, K., Fukushima, N., and Mori, Y. 2008. *Reference software for depth estimation and view synthesis.* ISO/IEC JTC1/SC29/WG11 MPEG, 20081, p.M15377.

[22] Technical Specification Group RAN. *E-UTRA: LTE RF system scenarios*. 3GPP Technical Report. TS 36.942, Dec. 2008-2009.