Application-level performance of cross-layer scheduling for social VR in 5G

Du, Z.; van den Berg, J.L.; Dimitrovski , T.; Litjens, R.

DOI
10.1109/WCNC55385.2023.10118957

Publication date
2023

Document Version
Final published version

Published in
2023 IEEE Wireless Communications and Networking Conference, WCNC 2023 - Proceedings

Application-level performance of cross-layer scheduling for social VR in 5G

Z. Du
Faculty of EEMCS
Delft University of Technology
Delft, The Netherlands
z.du@student.tudelft.nl

J.L. van den Berg
Faculty of EEMCS
University of Twente
Enschede, The Netherlands
j.l.vandenberg@utwente.nl

T. Dimitrovski
Department of Networks
TNO
The Hague, The Netherlands
toni.dimitrovski@tno.nl

R. Litjens
Department of Networks
TNO
The Hague, The Netherlands
Faculty of EEMCS
Delft University of Technology
Delft, The Netherlands
remco.litjens@tno.nl

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10118957

***Abstract*—Social VR aims at enabling people located at different places to communicate and interact with each other in a natural way. It poses extremely strong throughput and latency requirements on the underlying communication networks. This paper investigates the potential of using cross-layer design approaches for radio access scheduling in order to realize these challenging requirements in (beyond) 5G networks. In particular, we provide an in-depth simulation study of the performance/capacity gains that can be achieved by exploiting the end-to-end latency budget and/or video frame type as cross-layer information in the scheduling decisions, and show how the benefits depend on the actual social VR scenario. This study further reveals the importance of using application-level metrics such as PSNR or SSIM rather than traditional network-level metrics like the packet drop rate in the performance assessment.**

Keywords—Social VR, cross-layer scheduling, application-level performance, 5G

I. Introduction

One of the most challenging application classes targeted by (beyond-)5G networks is virtual reality (VR), in particular scenarios with multiple users located in different places who are able to interact smoothly with each other [1][2]. Such so-called social VR applications will decrease the need for travelling and hence save time and costs by enabling e.g. virtual meetings or other forms of collaborative working, remote exploitation of scarce expertise or skills in industry, and remote education and training. The COVID-19 pandemic happening in the past few years has also emphasized the relevance of these types of applications.

To realize the immersive experience of VR-based applications stringent service requirements have to be fulfilled by the network. In particular, besides the need for very high throughputs, the interactive, real-time nature of social VR asks for extremely low end-to-end (E2E) latencies [2][3]. Cross-layer solutions, where control information is shared among layers for increased adaptivity, are designed with an aim to meet these stringent requirements, see e.g. [4]. Examples of cross-layer approaches for supporting delivery of VR applications over mobile networks involve e.g. head movement or Field of View (FoV) prediction to reduce required throughputs [5], smart adaptation of video quality levels (e.g. image resolution) to the actual or predicted availability of network and/or computing resources [6] and advanced MAC- layer scheduling in Radio Access Networks (RAN) to minimize latency budget violations [7].

In this paper we focus on the performance and capacity gains that can be achieved by using cross-layer approaches for RAN scheduling. In particular, we investigate the potential of taking into account the E2E latency budget and/or video frame type as cross-layer information in the scheduling decisions. In *E2E latency-based scheduling* it is assumed that the remaining E2E latency budget of each individual packet of the involved VR sessions is available when it arrives at the base station. The main idea behind *frame type-based scheduling* is to exploit VR video frame type information encapsulated in the IP packets in order to prioritize the packets originating from I-frames, which are self-contained and independently encoded, over those from P-frames, which contain only changes in the image relative to that represented by the previous frame.

E2E latency- and frame type-based scheduling have been studied in existing literature, but mostly in contexts significantly different from the (social) VR over 5G scenario considered in this paper, see e.g. [7][8][9][10]. In particular, to our knowledge [7] is the only paper that studies frame type-based scheduling in the context of multi-user VR over 5G mobile networks. The frame type-based scheduler proposed in that paper will also be taken into account in our (much broader) study. The main contributions of our work are:

- An in-depth assessment of the performance and capacity gains that can be achieved by exploiting cross-layer information regarding the experienced/allowed E2E latency *and/or* video frame types in RAN scheduling, incl. a sensitivity analysis w.r.t. distinct scenario aspects.
- Showing the importance of using application-level rather than network-level metrics for the assessment of the performance impact of these schedulers.

The remainder of this paper is organized as follows. Section II describes the considered social VR setting. In Section III, the different (non-)cross-layer RAN schedulers are specified. Section IV then describes the key modelling aspects and defines the Key Performance Indicators (KPIs) used in the performance assessment. Extensive simulation results are then presented and discussed in Section V. Finally, in Section VI we summarize the key findings of our study.

II. Setting

A high-level view of the considered social VR setting is shown in Figure 1, depicting a virtual meeting with multiple participants at distinct locations. The central element of the system is the media processing unit or Multi-point Control Unit (MCU), which receives streams from all participants, combines them into a single scene and sends the encoded output of a viewport from that scene back to each user. It is

deployed in the cloud and sends a downlink packet stream covering the combined scene of participants via an indoor 5G base station (gNodeB) to the participants physically present in a meeting room. The packets traverse multiple routing hops of an IP network, representing the Internet, which introduces dispersion. At the end of the downstream chain we assume a meeting room with a table seating the physical participants wearing Head-Mounted Displays (HMDs), equipped with integrated User Equipment (UEs) to provide connectivity. Their mobility is limited to head movements.

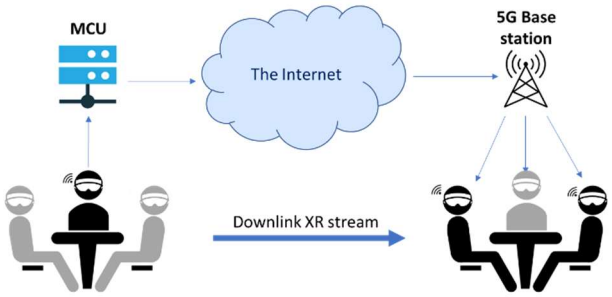

Figure 1: Social VR setting.

Each packet from the output stream of the MCU contains (fragments of) frames of type I and P, which are standard types used in video compression. A set of frames starting with an I-frame followed by the consecutive P-frames until the next I-frame is called a Group of Pictures (GoP). The size of the GoP is a configuration parameter of the encoder and sets the number of P-frames that will follow an I-frame. Using more P-frames in the encoding process makes the required bit rate for transfer lower, but introduces inter-frame dependencies which may lead to error propagation across frames.

In modern-day mobile networks the base station-based packet scheduler is a key real-time mechanism in charge of satisfying performance requirements, deciding on a millisecond timescale which QoS flow(s) is/are served on the available spectral resources. Scheduling decisions are traditionally influenced purely by locally known (RAN-oriented) input variables, including the flow-specific buffer status and an estimation of the instantaneous channel quality of the served UEs. Current scheduling implementations do *not* use application-level information like the (experienced/ required) E2E frame-level latency or whether a packet corresponds with an I- or P-frame. It is the chief objective of this study to propose and assess cross-layer scheduling solutions that *do* exploit such application-level information.

III. Cross-Layer Scheduling Solutions

As said, the packet scheduler in the base station is in charge of assigning spectral resources to active QoS flows. In 4G networks, the scheduling objective typically involves a targeted trade-off between resource efficiency and fairness among flows, commonly achieved by a suitable configuration of a wideband Proportional Fair (PF) scheduler [11]. In mixed-traffic 5G scenarios, the scheduler must be able to effectively handle not only throughput-hungry flows but also flows whose QoS requirements include a latency component. In the following we consecutively define the baseline (non-cross-layer) and cross-layer schedulers assessed in this study.

Baseline schedulers

A scheduler often applied in literature in such mixed-traffic scenarios is the *Modified-Largest Weighted Delay First (M-LWDF)* scheduler [12], which is an extension of the basic PF scheduler and assigns the available frequency carrier ('wideband scheduling') to flow i which maximizes

$$M_i^{\text{M-LWDF}}(t) = \frac{\log_{10}\delta_i}{\tau_{\text{RAN}}} W_{i,RAN}(t) \frac{R_i(t)}{\hat{R}_i(t-1)},$$

where δ denotes the maximum allowed PDR, τ_{RAN} denotes the RAN-oriented IP packet-level latency budget, $W_{i,RAN}(t)$ denotes flow i's experienced Head-Of-Line (HoL) IP packet latency experienced since its arrival in the base station buffer, $R_i(t)$ denotes the flow's instantaneously attainable bit rate in upcoming Transmission Time Interval (TTI) t, and $\hat{R}_i(t-1)$ denotes the exponentially smoothed bit rate it experienced so far. After scheduling for TTI t, $\hat{R}_i(t)$ is updated according to

$$\hat{R}_i(t) = \begin{cases} (1-\alpha)\hat{R}_i(t-1) + \alpha R_i(t) & \text{if scheduled} \\ (1-\alpha)\hat{R}_i(t-1) & \text{otherwise} \end{cases}$$

where smoothing parameter α is configured to strike a desired efficiency-vs-fairness trade-off.

The basic principle of the multiplicative component preceding the PF ratio in the M-LWDF scheduler is to increase a flow's scheduling priority as the latency of its HoL IP packet approaches the imposed latency budget. The scheduler further incorporates a packet dropping mechanism, removing (residual) IP packets from the buffer in case they cannot complete transmission within the latency budget.

Another latency-oriented scheduler considered in the study is the *Earliest Due Date* scheduler whose definition and (similar) evaluation results are not included here, due to lack of space, but are available in [13, Sections 3.2.4 and 4.2].

The described schedulers are noted to be purely RAN-oriented and hence can be executed entirely based on information which is de facto available to the base station, either through operator configuration (δ, τ_{RAN}), UE feedback ($R_i(t)$) or local processing ($\hat{R}_i(t-1)$, $W_i(t)$). We therefore classify these schedulers as non-cross-layer schedulers.

Cross-layer schedulers

As mentioned above, the goal of the paper is to assess to what extent cross-layer information can improve scheduling decisions leading to enhanced service quality and cell capacity. We consider two distinct scheduling enhancements, applied either in isolation or combined with the M-LDWF baseline scheduler, thereby establishing three cross-layer schedulers:

- In one enhancement we assume that the scheduler-incorporated *latency* aspects apply at the *E2E frame level*, rather than at the RAN-oriented IP packet level, as assumed in the baseline schedulers. Specifically, this means that the adapted scheduling metrics incorporate *(i)* τ_{E2E}, denoting the E2E frame-level latency budget; and *(ii)* $W_{i,E2E}(t)$, denoting the E2E latency experienced by the frame associated with flow i's HoL IP packet. Note that monitoring $W_{i,E2E}(t)$ requires that the scheduler is informed about the generation time of

frames at the source or, alternatively, about the accumulated latency experienced before arrival at its buffer. The associated packet dropping mechanism is triggered by a frame-level latency check.

- Another enhancement assumes the scheduler to apply *frame type-based scheduling*, assuming awareness of whether a buffered IP packet belongs to an I-frame or a P-frame, e.g. via deep packet inspection, and prioritizes (the from an application perspective more important) I- over P-frame-associated packets.

These enhancements yield three cross-layer schedulers, denoted M-LWDF-E2E, M-LWDF-FTS, M-LWDF-FTS/E2E, with 'FTS' referring to 'Frame Type-based Scheduling'. The scheduling metric of e.g. the M-LWDF-FTS/E2E scheme is given by

$$M_i^{\text{M-LWDF-FTS/E2E}}(t) = \varphi_i(t) \cdot \frac{\log_{10} \delta_i}{\tau_{\text{E2E}}} W_{i,E2E}(t) \frac{R_i(t)}{\hat{R}_i(t-1)},$$

where

$$\varphi_i(t) = \begin{cases} \varphi & \text{if } \text{HoL}_i(t) \in \text{I-frame} \\ 1 & \text{otherwise} \end{cases}$$

with prioritization weight $\varphi \geq 1$ and $\text{HoL}_i(t)$ denoting flow i's HoL packet at time t. In the experiments we use $\varphi = 5$. See [13, Section 3.2.4] for explicit definitions of all schedulers. As we will demonstrate, combined use of both types of cross-layer information is most promising in resource-efficiently provisioning for a targeted application-level quality level. The considered alternatives serve as relevant benchmarks, including the M-LWDF-FTS scheduler, which incorporates the cross-layer scheduling principle proposed in [7].

IV. Modelling

In this section we describe the key application, network and propagation modelling aspects and define the used KPIs.

Application/traffic model

The application is modelled by use of an actual video trace featuring a person 'in social VR mode' in order to best approximate the targeted setting. The video stream has been configured with a default bit rate of 100 Mbps (corresponding with an 'entry-level VR' session [14]), a frame rate of R_F = 30 fps [14] and a 10-frame GoP size [15], as visualized in Figure 2. We assume a default E2E latency budget of the social VR session of 50 ms. In the numerical experiments presented in section V we consider a range of settings for both the application bit rate and the E2E latency budget.

We have streamed the captured video frames through our network using GStreamer and subsequently captured the resulting IP packet trace with Wireshark, enabling us to obtain for each packet its size, the number of the frame the packet belongs to and the corresponding frame type (I or P). In practical implementations the video encoder releases the packets of a video frame in a burst, which is passed through a smoother before further releasing them into the network. To effectuate this in a controlled manner, we defined a *packet dispersion model*, configured by burstiness parameter β, in which the inter-packet release time is fixed at $\Delta t = (1 - \beta) / (N_{max} \cdot R_F)$, where N_{max} denotes the maximum of the number of packets per frame. This dispersion model ensures uniform inter-packet release times throughout the video stream, while avoiding overlap of packets belonging to different frames. In the numerical experiments we use $\beta = 0.6$.

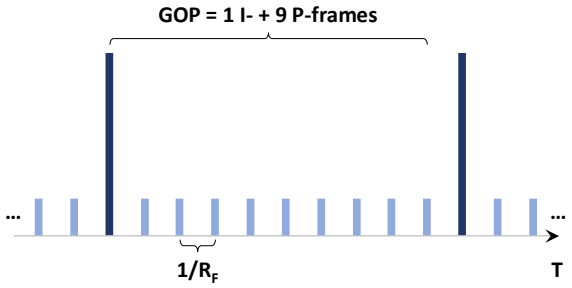

Figure 2: VR application traffic model.

Given multiple participants in the social VR session, each fed with an individual flow of packets, the relative timings of the I/P-frame releases feeding the participant-specific packet flows depends on when exactly the participants join the session and initiate the video stream. In our experiments we have considered both the worst case of perfectly synchronized I-frames, causing periodic spikes in aggregate traffic, and the best case of maximum asynchronization.

Network aspects: path from MCU to gNodeB

After the initial packet dispersion incurred at the video encoder, a flow's IP packets experience variable delays on their common internet *path from the MCU to the gNodeB*. As visualized in Figure 3, the path comprises M routers, with each router handling both the flow's tagged packets (red arrows) and a Poisson stream of background traffic (green arrows) in a first-in first-out fashion at a traffic handling rate of 1 Gb/s. We consider M = 5, 10 and 15 to model exemplary social VR sessions maintained between Amsterdam and Delft, Berlin and New York, respectively, with the corresponding M values based on traceroute experiments.

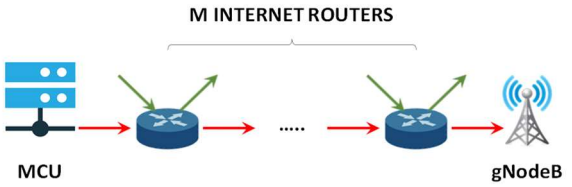

Figure 3: Network path from the MCU to the 5G base station.

Considering these three distinct geographical scenarios, each packet is modelled to experience a deterministic E2E *propagation delay* of 0.3, 3 and 30 ms, respectively, which is simply approximated based on the distance between the locations and assuming signals travel at the speed of light. The *queuing delay* experienced at a router by a tagged packet depends on the routers current buffer occupancy, while the packet further incurs a *transmission delay* determined by packet size and the router's traffic handling rate.

We assume independent and identically distributed Poisson streams of background packets at the M routers, characterized by an arrival rate of λ packets/second and a three-valued packet size (denoted s (in bytes)) distribution with $\mathrm{P}(s = 44) = 0.44$, $\mathrm{P}(s = 1300) = 0.19$ and $\mathrm{P}(s = 1500) = 0.37$, and hence an average packet size of about 821 bytes.

The arrival rate λ is set to 76103, 106610 or 129366 packets/second, corresponding with a respective background load on the routers of 50%, 70% and 85%.

Network aspects: radio interface

Considering the radio interface as the final hop on the downstream E2E path, we consider a single ceiling-mounted indoor base station featuring a simple omni-directional antenna. The base station is assigned a 150 MHz (default value) TDD (Time Division Duplexing) carrier in the 3.5 GHz band, with a 4:1 downlink/uplink slot ratio and configured with OFDM (Orthogonal Frequency Division Multiplexing) numerology 2. Active UEs are assumed to provide CSI (Channel State Information) feedback every 5 TTIs. Based on these reports, the gNodeB estimates for each UE i the SINR (Signal to Interference plus Noise Ratio), maps this to an MCS (Modulation and Coding Scheme) and derives the correspondingly attainable bit rate $R_i(t)$ for the upcoming TTI t. These $R_i(t)$'s feed into the scheduler, as discussed above. A biased coin is flipped based on the applied MCS and the experienced SINR to determine whether the transport block is successful or needs to be retransmitted.

For the considered office scenario we adopt the indoor Line-of-Sight *propagation* model from [3], covering path loss, shadowing and multipath fading effects, and utilize the model implementation provided by QuaDriGa [16]. Affecting multipath fading, head movement in terms of change of direction and randomized wobbling is modeled as in [17].

KPIs

In the quantitative assessment of the different schedulers three distinct KPIs are used, viz. the network-level PDR metric and the application-level PSNR and SSIM metrics.

The *PDR* (Packet Drop Rate) is a network-level metric, given by the fraction of IP packets that is dropped, either by the base station scheduler or by the application layer at the UE side, in case a packet delay exceeds the E2E latency budget of the corresponding video frame.

The *PSNR* (Peak Signal to Noise Ratio) [18] is a full-reference application-level metric and hence based on a direct comparison of the original and received video streams. The PSNR is defined as $10\ log_{10}(R^2/MSE)$, where R denotes the maximum fluctuation in the input image data type (h.l. R = 255) and *MSE* denotes the Mean Squared Error of the pixel- and color-specific power differences between the original and the received images. The *average PSNR* of a video stream is given by the average of the image/frame-level PSNRs. According to [19], an average PSNR exceeding 31 dB indicates good video quality.

The *SSIM* (Structural SIMilarity) [20][21] is another full-reference application-level metric, considering changes in structural information to estimate the image degradation with a value ranging between 0 (poor) and 1 (perfect). See [21] for a detailed description of the SSIM measurement system. The *average SSIM* of a video stream is given by the average of the image/frame-level SSIMs. According to [21], an average SSIM exceeding 0.95 indicates good video quality.

V. Simulation Scenarios & Results

The simulation scenarios are defined by a set of operator-controlled or environment/service-specific parameters. We listed the most important scenario parameters in Table 1, with ranges indicated and the assumed default value italicized.

Table 1: Scenario parameters.

Parameter	Values
Application bit rate	50-200 (*100*) Mbps
E2E latency budget (τ_{E2E})	25, *50*, 100 ms
RAN latency budget (τ_{RAN})	1, 5, 10, *20*, 30, 40, 50 ms
Carrier bandwidth	100, *150*, 200 MHz
Number of UEs	1-8 (*4*)
Background traffic load	50, *70*, 85%
Number of network hops	5, *10*, 15
Synchronization of video traffic stream encoding	*maximum asynchronization*, perfect synchronization

In order to make informed decisions on how PDR would impact a service, we used the guideline that acceptable network PDR without too much degradation is around 5% [22]. To aid in selecting a default setting for τ_{RAN}, Figure 4 shows simulation results for the PDR versus different τ_{RAN} options, assuming the non-cross-layer M-LWDF scheduler and considering different number of network hops and background traffic loads. All curves show the same basic pattern: for *very small* τ_{RAN} values many data packets are dropped pro-actively (but possibly unnecessarily) by the base station scheduler; in contrast, for *very large* τ_{RAN} values there is hardly any pro-active packet dropping and virtually all packets are actually transmitted, which in turn imposes a (wastefully) high cell load, consequently high end-to-end latencies and excessive dropping by the application at the UE. Discarding the 15-hop scenarios (showing poor performance anyway as the propagation delay already consumes a big chunk (30 ms) of the τ_{E2E} = 50 ms) and concentrating on the remaining ones, a choice of 20 ms seems to be a good setting for τ_{RAN}, which is therefore selected as our default value.

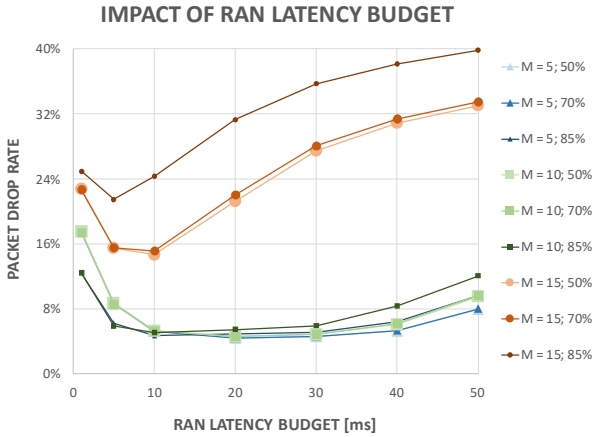

Figure 4: Impact of τ_{RAN} on the PDR for the baseline (non-cross-layer) M-LWDF scheduler.

A detailed analysis of all the scenario parameters and their impact can be found in [13, Section 4.2]. Indeed, using frame-type information in scheduling decisions does not seem to give notable performance improvements over baseline scheduling in view of PDR. In fact, it usually performs worse for that metric. Similar can be said for using τ_{E2E} as opposed to the τ_{RAN} in scheduling decisions, although in some cases it does perform better. Furthermore, tuning the τ_{RAN} budget in

RAN latency-based schedulers makes them come very close in performance to E2E latency-based schedulers making it a seemingly suitable solution for an operator that aims to optimize a social VR application with the current baseline schedulers. We stress however that many scenario parameters cannot be controlled and are unknown to an operator which makes the process difficult. The main consequence of using frame-type information in the scheduling is the shift of the packet loss heavily towards the P-frames. This means a different kind of impact on the service.

In order to quantify this impact on the application level, we used the PSNR and SSIM values for each frame after it is decoded, and calculated the average values per each simulation run. Figure 5 depicts these values for all runs including all variations of the scenario parameters given in Table 1. The first thing to observe from the figure is the exponential relation between the PDR and PSNR. The consequence of this is that when looking at the PDR on the network level, a small increase can either mean no difference on the application or a very large impact depending on the range. Hence, in order to more accurately grade the service quality and the benefits/drawbacks from the cross-layer scheduling enhancements, application level metrics are the most relevant rather than network level because they may not be linear or may even be bad proxies of actual application-level performance.

In this particular case, even for high levels of PDR between 20-30% which usually entails a bad service, the PSNR values for a M-LWDF-FTS scheduler are above the 31 dB threshold meaning that the most relevant information for the decoder did get through the network. The reason for this is the fact that I-frames are the pillars of that information so

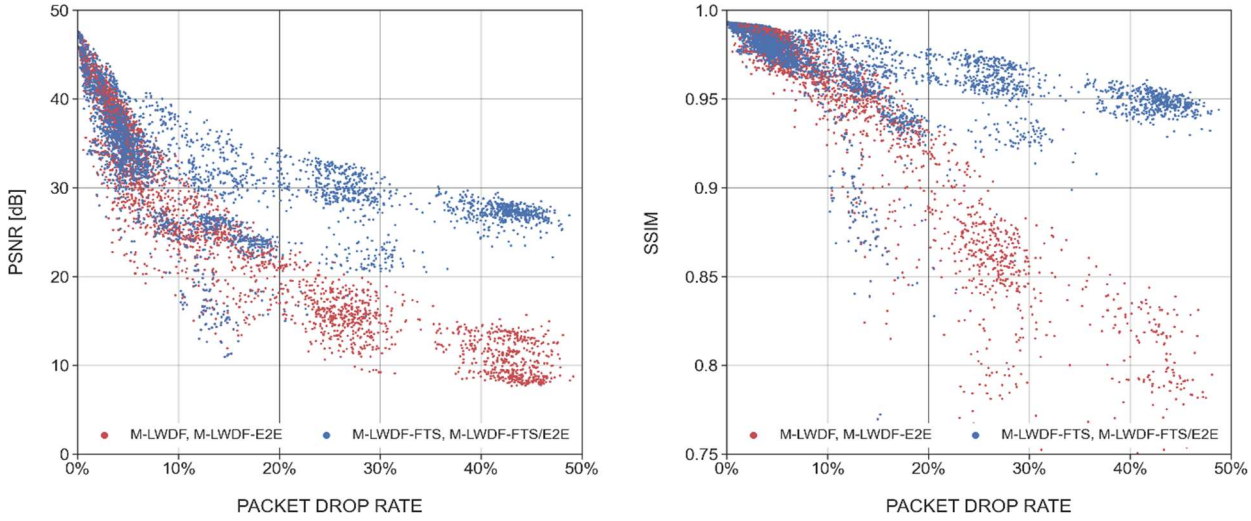

Figure 5: Relation between the application-level (PSNR, SSIM) and network-level (PDR) performance metrics for the non-frame type-based (M-LWDF, M-LWDF-E2E) and the frame type-based (M-LWDF-FTS, M-LWDF-FTS/E2E) schedulers.

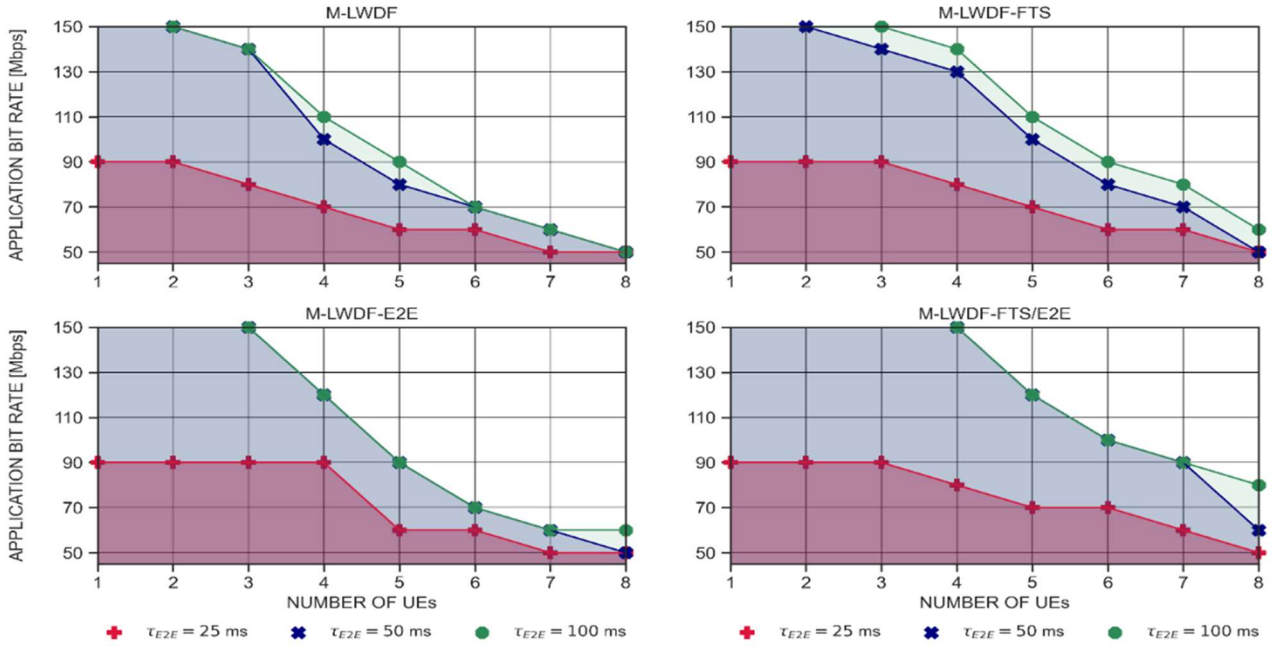

Figure 6: Feasible scenarios for different load parameters.

preserving them can keep the application impact lower. Effectively, this becomes a multi-objective optimization because it decreases the impact of the PDR and increases the PSNR. The combination of preserving I-frames with proportional fairness goals seem to make the best use of the radio resources when they are scarce. This can be concluded from the fact that the blue scattered dots (FTS schedulers) tend to be in the higher levels of PSNR and SSIM than the red scattered dots (non-FTS schedulers) depicted on Figure 5.

Finally, the relative gains from the two scheduling enhancements can be expressed in the ability of the system to deliver good service quality (PSNR ≥ 31 dB) for more load scenarios by varying the relevant parameters from Table 1, i.e. the application bit rate and the number of UEs. Figure 6 depicts this for the M-LWDF scheduler, where the attainable gains for τ_{E2E} = 25 ms are very modest due to having much less space in the latency budget, and are biggest for the 100ms cases as this space gets larger. If we look at the more realistic cases with τ_{E2E} = 50 ms and 100 Mbps application bit rate, when compared to the baseline (upper left plot), the percentage gains from using τ_{E2E} in scheduling (lower left) is 12.5%, from using frame-type information (upper right) its 25% and from using both (lower right) its 50% meaning the addition of two more UEs. In terms of Mbps with 4 UEs the gains are similar, with 20%, 30% and 50% respectively meaning the addition of 50 Mbps per UE.

VI. Conclusions

We have carried out an extensive simulation study in order to analyze the potential of different cross-layer RAN scheduling approaches for supporting highly demanding social VR applications in (beyond) 5G mobile networks. Our study has focused on exploiting E2E latency budget and video frame type (I- or P-frame) as cross-layer information in the scheduling decisions. In our simulations we used video traffic traces representative for the social VR use case, a wide-area network model to mimic IP packet delays on Internet paths and realistic social VR user behavior and radio propagation models. The simulation results show that if the PDR is used as performance metric, the added value of the cross-layer schedulers compared to benchmark (non-cross-layer) schedulers is very limited. However, considering application-level metrics like the PSNR or the SSIM, involving cross-layer information in the packet scheduling decisions leads to a significant improvement in network efficiency and/or social VR application quality. In particular, for realistic application bit rate and E2E latency requirements we found a RAN efficiency improvement of 50% in terms of the number of social VR users that can be supported. In future research we will investigate benefits of using additional cross-layer information in scheduling decisions, e.g. the number/type of buffered packets, in order to further reduce drop of I-frame packets.

Acknowledgment

The authors appreciate the contributions of Jan Willem Kleinrouweler and Rick Hindriks in modelling social VR application aspects, and of João Morais in modelling the 5G radio interface. This research was done when Z. Du was with the Department of Networks, TNO, The Netherlands.

References

[1] 5G PPP, *'5G innovations for new business opportunities'*, white paper, see https://5g-ppp.eu/wp-content/uploads/2017/01/5GPPP-brochure-MWC17.pdf, 2017.

[2] M. Latva-aho and K. Leppänen (eds.), *'Key drivers and research challenges for 6G ubiquitous wireless intelligence'*, white paper, University of Oulu, Finland, 2019.

[3] 3GPP, *'Extended Reality (XR) in 5G'*, TR26.928, v16.0.0, 2020 .

[4] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters and F. De Turck, 'Toward truly immersive holographic-type communication: challenges and solutions', *IEEE Communications Magazine*, vol. 58, no. 1, 2020.

[5] J. van der Hooft M. T. Vega, S. Petrangeli, T. Wauters and F. De Turck, 'Optimizing adaptive tile-based virtual reality video streaming', *Proceedings of IM '19*, Washington DC, USA, 2019.

[6] X. Hou, S. Dey, J. Zhang and M. Budagavi, 'Predictive adaptive streaming to enable mobile 360-degree and VR experiences', *IEEE Transactions on Multimedia*, vol. 23, 2020.

[7] M. Huang and X. Zhang, 'MAC scheduling for multiuser wireless Virtual Reality in 5G MIMO-OFDM systems', *Proceedings of ICC Workshops '18*, Kansas City, USA, 2018.

[8] Y. Zhang, Y. Zhang, S. Qin, B. Li and Z. He, 'Delay-bounded priority-driven resource allocation for video transmission over multihop networks', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 7, 2014.

[9] W. Kim, H. Joo, K. J. An, I. Lee and H. Song, 'Urgency-based packet scheduling and routing algorithms for delay-sensitive data over MANETs', *Wireless Networks*, vol. 19, no. 7, 2013.

[10] T. Ojanperä, M. Uitto and J. Vehkaperä, 'QoE-based management of medical video transmission in wireless networks', *Proceedings of NOMS '14*, Krakow, Poland, 2014.

[11] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana and A. Viterbi, 'CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users', *IEEE Communications Magazine*, vol. 38, no. 7, 2000.

[12] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia and P. Camarda, 'Downlink packet scheduling in LTE cellular networks: key design issues and a survey', *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, 2013.

[13] Z. Du, *'Cross-layer optimization of MAC scheduling for multi-user Virtual Reality over 5G'*, MSc thesis, Delft University of Technology, 2022.

[14] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran and M.D. Silva, 'VR is on the edge: how to deliver 360° videos in mobile networks', *Proceedings of VR/AR Network '17*, Los Angeles, USA, 2017

[15] H. Yao, R. Ni and Y. Zhao, 'Double compression detection for H.264 videos with adaptive GOP structure', *Multimedia Tools and Applications*, vol. 79, 2019.

[16] Fraunhofer, *'Quadriga – The next generation radio channel model'*, https://quadriga-channel-model.de, 2020.

[17] J. Morais, S. Braam, R. Litjens, S. Kizhakkekundil and J.L. van den Berg, 'Performance modelling and assessment for social VR conference applications in 5G radio networks', *Proceedings of WiMob '21*, virtual conference, 2021.

[18] Q. Huynh-Thu and M. Ghanbari, 'The accuracy of PSNR in predicting video quality for different video scenes and frame rates', *Telecommunication Systems*, vol. 49, no. 1, 2012.

[19] K. Bouraqia, E. Sabir, M. Sadik and L. Ladid, 'Quality of experience for streaming services: measurements, challenges and insights', *IEEE Access*, vol. 8, 2020.

[20] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, 'Image quality assessment: from error visibility to structural similarity', *Transactions on Image Processing*, vol. 13, no. 4, 2004.

[21] J. R. Flynn, S. Ward, J. Abich and D. Poole, 'Image quality assessment using the SSIM and the just noticeable difference paradigm', *Proceedings of EPCE '13*, Las Vegas, USA, 2013.

[22] A. O. Adeyemi-Ejeye, M. Alreshoodi, L. Al-Jobouri, M. Fleury, and J. Woods, 'Packet loss visibility across SD, HD, 3D, and UHD video streams', *Journal of Visual Communication and Image Representation*, vol. 45, 2017.