Enabling Virtual Reality for the Tactile Internet: Hurdles and Opportunities

Maria Torres Vega*, Taha Mehmli†, Jeroen van der Hooft*, Tim Wauters*, Filip De Turck*
* IDLab, Department of Information Technology, Ghent University - imec
Email: maria.torresvega, jeroen.vanderhooft, tim.wauters, filip.deturck@ugent.be
† École Nationale D'Ingénieurs de Tunis
Email: mehmlitaha@gmail.com

***Abstract*—The Tactile Internet envisions the real-time, full-fledged control of remote infrastructures by means of tactile sensors (haptics) and aided by audiovisual immersive inputs. End-to-end delays of at most 5 ms (minimum detectable by the human eye) and ultra-high-speed transmissions will be needed by services such as telehealth or Industry 4.0. To comply with these requirements, both networks and applications need to be dramatically improved. On the one hand, in the last years a massive effort is being put onto improving the current network infrastructures. Examples of this are the 5G paradigm or the advances on mechanisms of network precision control using network slicing, Software-defined networking solutions or novel service oriented protocols. On the other hand, Virtual Reality (VR) video streaming, which is the best-suited technology to provide the immersive visual feed of the Tactile Internet, is still far from the envisioned levels of real-timeliness and quality. The aim of this paper is to pinpoint the open challenges to enable VR video streaming technologies for the Tactile Internet. This paper first provides a thorough analysis of the state-of-the-art on VR video streaming technologies both, theoretically and by means of an experimental demonstrator. Based on this, we present the different research areas and key challenges within. In addition, possible solutions and research directions are introduced. We believe this work provides the means to open new opportunities for research not only within the challenging VR arena, but also in the field of management and control of wireless networks.**

***Index Terms*—Virtual Reality Video Streaming, Tactile Internet, Haptics**

I. Introduction

The Tactile Internet[1] envisions real-time monitoring, management and control of remotely located infrastructure [1]. The new paradigm aims to enable a plethora of new applications in contexts such as telehealth, Industry 4.0 (i4.0) or immersive online gaming [2]. This real-time remote control of machinery (*e.g.*, industrial machines or surgical devices) is to be achieved with the help of tactile sensors (haptics), whose kinesthetic feedback aids the human operator to drive the remote instrument. Audio and immersive visual feeds, *e.g*, Virtual Reality (VR) video, are proposed to further assist the interaction between the machinery and the operator. The integration of the visual and tactile real-time feeds requires novel infrastructures and services able to cope with the stringent requirements of the Tactile Internet. In fact, the new paradigm proposes end-to-end latencies (round-trip) between 1 and 5 ms (5 ms is the maximum latency that goes undetectable by the human eye). In addition, to cope with the stringent requirements of the visual feeds in terms of bandwidth, ultra-high-speed transmissions are to be expected. To achieve these characteristics, the Tactile Internet proposes a novel end-to-end architecture (Figure 1) [2].

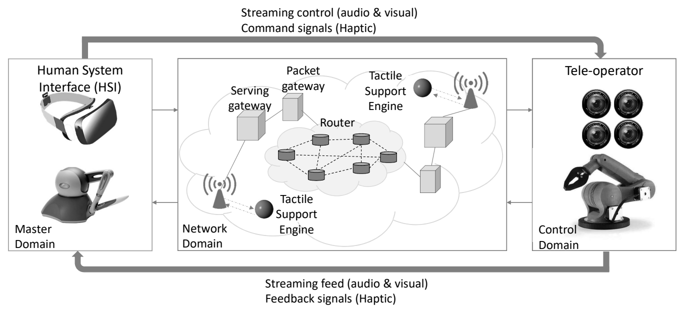

Fig. 1. End-to-end architecture of the Tactile Internet.

Three domains are defined: the master, the network, and the slave. In the master domain, an operator (typically a human) manages the Human System Interface (HSI), which in turn controls the remote machinery installed at the slave domain. The HSI consists of a set of haptic devices aided by an immersive visual feed. Haptics are instruments that involve physical contact between the computer (or machinery installed at the slave domain) and the user, usually through an input/output device. The purpose of haptic or kinesthetic communication is to recreate the sense of touch by applying forces, vibrations, or motions to the user [3], in order to enhance the remote control of machines and devices (telerobotics). Examples of such devices are a joystick, data gloves or the futuristic Tesla suit[2]. To provide the immersive visual feed required, the HSI includes a VR Head-Mounted Device (HMD), to which a real-time video feed from the slave is streamed. The slave domain (at the opposite end of the chain)

Jeroen van der Hooft is funded by grant of the Agency for Innovation by Science and Technology in Flanders (VLAIO). Maria Torres Vega is funded by grant of the Research Foundation Flanders (FWO) (grant number 12W4819N). This research was performed partially within the project G025615N Optimized source coding for multiple terminals in self-organizing networks from the fund for Scientific Research - Flanders (FWO-V).

[1]https://www.itu.int/en/ITU-T/techwatch/Pages/tactile-internet.aspx

[2]https://teslasuit.io/

978-3-903176-14-0

consists of a tele-operator (a controlled robot, in most of the cases) and the means to record and stream the visual feed (typically a set of cameras providing a full 360° view). By means of the VR feed and kinesthetic feedback, the operator is fully immersed in the remote environment [2], [4]. Between master and slave, the purpose of the network domain is to couple the human to the remote environment [4]. Even if the underlying network were to achieve near-real-time communication, the Tactile Internet must deal with the fundamental limitation set by the finite speed of light [1]. To overcome this, the Tactile Internet is envisioned to support an hybrid composition of machine and human actuation mixing real tactile action with intelligence-based prediction close to the edges of the network (dubbed Tactile Support Engines in Figure 1). These engines are to facilitate predictive caching as well as interpolation/extrapolation of human actions, by means of artificial intelligence solutions. In order to cope with the Tactile Internet, both network infrastructure and applications need to suffer a complete change of structure and working principles.

The current challenge of wireless networks is to reduce the end-to-end latency to the lowest limit (set by the speed of light) [2]. There are currently several research lines within the 5G paradigm [4] working towards designing core wireless networks with 1 ms round-trip latency. Examples of this are the applications of new wireless beyond Wi-Fi, such as the indoor 10 Gbps per user optical wireless beam-steering concept [5], or the 60 GHz based 100 Gbps femto-cells [6]. In addition, a plethora of network optimization algorithms and solutions have appeared in the last years. Network slicing approaches have been proposed for 5G [7]. Moreover, novel protocol stacks are current being developed. Such is the case of BPP a protocol stack focused on precise service guarantees [8].

Integrating VR video streaming within the Tactile Internet imposes necessary improvements on the transport technology. For example, current over-the-top (OTT) VR streaming services encode their 360° videos at ultra-high definitions (4K). Therefore, they require very high transmission bandwidths (50 Mbps) that are not always available in wireless networks and are not easy to process by lightweight mobile devices. Moreover, due to the streaming protocols, VR streaming suffers from large end-to-end latencies (ranging between 100 ms and seconds). Although not detrimental in video-on-demand (VoD) scenarios, these shortcomings are not compatible with real-timeliness. In addition, the monitoring and management of the perception of the users [4] and how to integrate the visual feed with the tactile feedback are still open issues.

Due to the early stages of development of VR video streaming technologies, these and other challenges are still open for discussion. The aim of this paper is to pinpoint the open challenges of VR video streaming in order to unlock its full potential within the Tactile Internet. In this paper, we first provide a thorough theoretical and experimental comparison of the state-of-the-art on video streaming solutions applied to the VR video case (Sections II and III). Then, the different research areas are presented, challenges are discussed and possible research solutions are introduced (Section IV).

II. VR VIDEO STREAMING PROTOCOLS

OTT multimedia streaming provides the best solution to transmit video (2D, 3D, or VR) over the Internet, when no dedicated network is available. Video streaming services are categorized as either using the Transmission Control Protocol (TCP) or the User Datagram Protocol (UDP). While streaming solutions which prioritize quality over latency tend to select TCP, solutions in which real-timeliness is the key optimization parameter opt for UDP.

Among the TCP-based solutions, RTMP (Real-time Media Protocol) and HTTP Adaptive Streaming (HAS) [9] are the most adopted ones. Both approaches try to optimize the streaming by means of splitting and sensing of the network. RTMP, on the one hand, splits streams into fragments in order to deliver streams smoothly and transmit as much information as possible. The size of the fragments is negotiated dynamically between the client and server and fragments from different streams may then be interleaved, and multiplexed over a single connection. The interleaving and multiplexing is performed at the packet level, with RTMP packets across several different active channels being interleaved, ensuring that each channel meets all Quality of Service (QoS) requirements, bandwidth and latency. HAS, on the other hand, performs segmentation on the application level. It encodes the streams at different quality levels, and temporarily splits them into segments of predefined duration. When the streaming takes place, the client decides the quality on which to request the next segment. Requests and transmissions are performed using HTTP (Hypertext Transfer Protocol). In an attempt to optimize the bandwidth requirements of ultra-high bandwidth applications (such as VR), HAS also allows to split the video streams not only in time but also in space (tiles). Thus, the HAS client is able to decide which areas of the scene presented in the segment to download at higher qualities. This feature becomes very useful for the case of VR video streaming, where only a small part of the scene (*i.e.*, the area covered by the HMD or viewport) is required at the highest quality at every instant [10]–[12].

Traditionally, streaming has been performed almost exclusively by TCP-based solutions where quality is prioritized over latency. However, this type of streaming induces undesired additional latency due to the constant packet acknowledgment, and are thus, not suitable for real-time transmissions. Contrary to TCP, UDP does not require packets to be independently acknowledged. As such, UDP solutions are better suited to provide the low-latency requirements. However, these solutions have shown problems to maintain quality levels (which is necessary for immersiveness in VR). Approaches such as QUIC (Quick UDP Internet Connections) [13] and WebRTC (Web Real-Time Communication) [14] are improving the performance of UDP-based solutions. QUIC is an experimental transport layer protocol designed by Google, which supports a set of multiplexed connections between two end-points over UDP. QUIC's main goal is to improve perceived

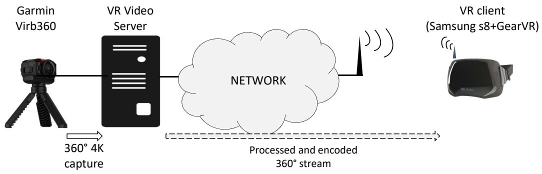

Fig. 2. 360° video experimental testbed.

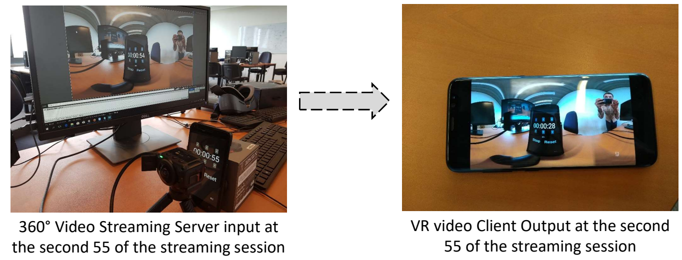

Fig. 3. Latency measurement for the case of using the xSplit Broadcaster combined with YouTube.

performance of connection-oriented web applications that are currently using TCP. WebRTC is a free, open-source project that provides web browsers and mobile applications with real-time communication (RTC) via simple application programming interfaces (APIs). This enables audio and video communication to work inside web pages by means of direct peer-to-peer communication, eliminating the need to install plugins or download native apps. Its mission is to enable rich, high-quality RTP applications to be developed for the browser, mobile platforms, and Internet of Things devices, and allow them all to communicate via a common set of protocols.

III. Experimental evaluation of VR video streaming services

In this section we aim to provide a quantitative analysis of the delay induced by current available solutions for video streaming services. Therefore, we assess their applicability to the VR video case and try to understand how far they are from the stringent requirements of the Tactile Internet.

To perform such quantitative evaluation, we set up a real-time VR streaming testbed (Figure 2). A 360° camera (Garmin Virb 360[3]) is connected via HDMI to a video capture card (Blackmagic Design Decklink Mini Recorder 4K[4]) installed on a computer running Windows 10 (i7 , 64Gb RAM, 12 cores , Nvidia Geforce 780ti). The card captures footage from the camera at 4K quality (3840×2160), using a frame rate of 29.97fps and a pixel format of 10 bits RGB. The computer acts as the VR server of the video streaming service to which a mobile VR client is connected (Samsung S8[5] and the Samsung GearVR goggles[6]). Both VR server and client are connected within the same WiFi network.

In order to evaluate the latency induced by the processing, streaming application protocol and network, a timer is set in front of the camera. The footage is captured and streamed to the device through the network. Then, the end-to-end latency is measured by direct subtraction of the current time and the current received value at the end device. Figure 3 illustrates this procedure.

Following this procedure, we tested a plethora of currently used streaming services, from which we picked six representative cases (Table I). We selected two different capturing and broadcasting engines, xSplit[7] and Wowza[8] due to their

TABLE I
Comparison among different commercial video streaming solutions for the case of VR streaming.

Solution	Prot.	Enc.	Resol.	Bitrate [Mbps]	Delay [s]
xSplit & YouTube (xSplit Client)	RTMP	H.264	1440p	17.48	26
xSplit & YouTube (Browser)	RTMP &QUIC	H.264	1440p	17.48	11
XSplit &Facebook (xSplit Client)	RTMP	H.264	720p	4	16
XSplit &Facebook (Browser)	RTMP &QUIC	H.264	720p	4	11
Wowza (4K)	RTMP	H.265	4K	20	10
Wowza (SD)	RTMP	H.264	480p	2	3

broad usage in video streaming services. xSplit first downsizes the video to 1440p before inputting it to well-known Internet streaming agents (such as Facebook[9] or YouTube[10]) or directly to the browser (Chrome). Youtube allows the streaming of live video at 1440p, while, Facebook resizes the video to 720p and limits the transmission speed to 4Mbps. Wowza, on the other hand, provides an end-to-end solution, streaming from their software-cloud server to the client at 4K quality. The tested streaming solutions base their transmissions on RTMP. Only the two solutions which stream to the browser combine RTMP with QUIC for the last part of the streaming (*i.e.*, streaming from xSplit to the Facebook/YouTube server using RTMP and streaming within the browser using QUIC).

These six solutions provide a full range of latencies for real-time streaming. As can be observed Table I, nearly all streaming solutions show at least 10 s of latency. Only in the case of streaming at 480p for the Wowza engine provides a solution of below 5 s. The reason why Wowza performs better than all other solutions comes from the fact that they provide full end-to-end solutions with their own control paths and cloud-based solutions. In addition, it is worth noticing the benefits of using browser-based solutions, in which the delays are reduced by at least 5 s from their full RTMP versions.

[3]https://buy.garmin.com/en-US/US/p/562010#specs
[4]https://www.blackmagicdesign.com/products/decklink/W-DLK-33
[5]https://www.samsung.com/global/galaxy/galaxy-s8/
[6]https://www.samsung.com/global/galaxy/gear-vr/
[7]https://www.xsplit.com/#broadcaster
[8]https://www.wowza.com/products/streaming-engine
[9]https://www.facebook.com
[10]https://www.youtube.com

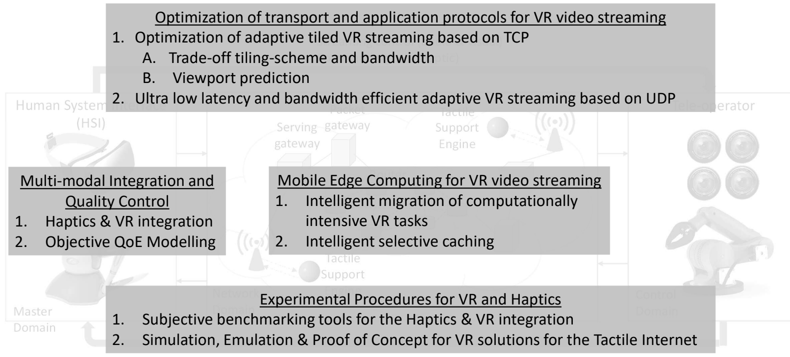

Fig. 4. VR open research challenges.

However, even with such improvement, the solution are still very far from the requirements of the Tactile Internet.

IV. Open Research challenges for VR video streaming services

In order to achieve real-timeliness and quality, current video streaming technologies have still to undergo a whole renovation process. These updates are not only required on the streaming protocols, but also on other aspects such as quality assessment and integration with the tactile feedback. Based on the state-of-the-art and our own experimental early results (shown in the previous two Sections), we have identified four different areas of research: (1) optimization of streaming protocols; (2) Mobile Edge Computing for VR streaming; (3) multi-modal integration and quality control and; (4) experimental evaluations and subjective benchmarking. Figure 4 shows the four different areas of research placed on top of the Tactile Internet end-to-end element to which it refers to. The next four subsections provide details on state-of-the-art, future directions and proposed solutions for each of these areas.

A. Optimization of transport and application protocols

A VR video contains a full 360° panoramic view, regardless of the fact that only a fraction of it, the viewport, is visible to the user at any given instant. This means that the adaptations at the application level (segmentation and tiling) proposed by the HAS approaches for VR [10] are better suited for bandwidth usage optimization than the packet level optimizations proposed by RTMP.

However, HAS for VR videos is still a relatively new field of research (first papers date back to 2016 [11]). Open challenges still exist and can be classified in three categories: (1) the trade-off between tiling granularity and bandwidth (*i.e.*, how the number of tiles affect the bandwidth requirements); (2) the viewport prediction (i.e., how to continuously predict where the user will look next); and (3) the performance in terms of minimization of the latency, maximization of the bandwidth utilization and quality (*i.e.*, how to optimize the streaming process to reduce the end-to-end latency while optimizing the bandwidth requirements).

First, there is no general set of rules to provide guidance on the selection of a tiling scheme depending on the conditions and bandwidth requirements. Approaches in the state-of-the-art have chosen different tiling schemes to provide their evaluation. Examples of these are the six tiles of [10] or the 32 tiles of [11], [12]. This lack of a standard makes it difficult to benchmark new solutions. Given the mathematical essence of such trade-off, a possible first step could be to formalize the trade-off by means of mathematical modeling techniques such as Linear Programming and Genetic algorithms. In a second step, heuristic solutions would help on real-time decisions regarding the best combination of tiling scheme for the current conditions, videos and users.

Second, viewport prediction is a very sensitive task that affects the users' perception in unexpected ways. Errors on the prediction of the current viewport, lead to partial or full degradation of the perception, even if the network conditions should guarantee the users' Quality of Experience (QoE) [15]. It has been shown that performing such predictions can reduce bandwidth utilization up to 70% [12]. However, the current approaches lack generality, as these are only tested with very limited network conditions, users, and video content. Thus, the design of a generic viewport prediction algorithm valid for all conditions and videos is an important open research topic. In addition, the viewport prediction scheme needs to be accurate and fast. In order to fulfill all these characteristics, we hypothesize that the position where the users look in the next instant not only depends on the movement speed (which can be measured by the HMDs), but also on the type of content that the user is visualizing. However, this relation is highly non-linear. In this track, novel machine learning techniques (both supervised and reinforcement learning) have be brought to play to model this relationship.

Finally, as we introduced in the previous Section, the TCP essence of HAS imposes unwanted additional delays. Thus, there is a need to reduce the latency as much as possible. Some advances towards minimizing the latency within TCP suggest the use of HTTP/2 [10], or other TCP based approaches such as TCP/BBR [16] and TCP datacenter (DCTCP) [17].

As we also showed, UDP can provide the very low-overhead alternative [2]. However, what these protocols gain in real-timeliness, they lose in bandwidth and quality delivery assurance: since they do not possess control mechanisms, they are heavily affected by losses in the networks. This circumstance can be easily countered for the case of 2D multimedia content (for example, WebRTC [14] adapts the real-time encoding bitrate of the videos in function of the sensed networks). However, the losses can be dramatic for the transmissions of the ultra-high bandwidths required by VR content. Due to this fact, one possible solution would be to develop a transport/application protocol integrating the tile-based adaptive streaming techniques of HAS on to RTP/UDP. This possible path has been shown by Zhao et al. [18] for the case of 2D video streaming using WebRTC. However, the case of VR is still open for exploration.

B. Mobile Edge Computing within the Tactile Internet

As introduced in Section I, the network domain of the Tactile Internet envisions to place intelligent units at the edges of the network to improve latency (*i.e.*, Tactile Support Engines). Live VR video streaming could benefit from such intelligent edge infrastructure. We identify two possible applications: (1) migration of computationally intensive tasks from both the server and the client to the edges; and (2) intelligent selective pre-caching and pre-fetching.

Actions such as the viewport prediction, the segment/tiles quality assignment or the synchronization of the kinesthetic feedback can become computationally too intensive for the lightweight VR client devices to cope with. Furthermore, the real-time encoding process of ultra high resolution becomes very computationally intensive, even for the video capturing server. In such cases, being able to partially or totally migrate the tasks to the more resourceful cloud or fog infrastructure would increase the computational capacity of low-cost devices as well as saving battery.

Proactive caching allows to cache strategic content based on fine-grained traffic and user predictions [4]. Information such as the users' location, mobility patterns, and social ties can be further exploited, especially when context information is sparse. Storage will play a crucial role in VR where, for instance, upon the arrival of a task request, the network/server needs to rapidly take a decision on whether to store the object in the case the same request comes in the near future. This type of pre-caching could be very useful for envisioned cases such as the remote control of a robotic arm able to control certain actions in a factory, with an installed fixed camera feed. Actions performed by the operator will often follow a determined pattern. Thus, by predicting the location where the operator will be looking next (based on current position and patterns), the system could prefetch and cache the next content required, thus reducing latency. Deciding which content/views to prefetch and cache on the edges could be performed by means of user profiling techniques and pattern recognition.

C. Multi-modal Integration and Quality Control

Human reaction times depend heavily on the source and sensing input [2]. While only 1 ms is required for the brain to receive a tactile signal, it takes 10 ms to understand visual information, and up to 100 ms to decode audio [4]. This situation brings two fundamental challenges with it: (1) the real-time feedback of the different systems needs to be adapted to the different speeds, thus synchronizing the inputs; (2) It is necessary to continuously monitor the quality perceived by the users of the applications, thus understanding desynchronizations and problems derived from wrong performance of the immersive environment.

The human brain integrates sensor modalities with different requirements in terms of sampling, delays, etc. [19]. This forces the need for multiplexing schemes, capable of exploiting priorities, as well as temporal integration of different modalities. Some initial research efforts aim at addressing this challenge [19], by means of adaptive multiplexing of visual and haptic data. These approaches have provided promising results. However, the performance under dynamically changing environments as in wireless networks, and in particular their error-resiliency to packet losses, has not been investigated. In addition, they have not shown the integration of VR content, which requires a more complex synchronization scheme than 2D videos. Further developments in this area are required to achieve the truly immersive steering and control envisioned.

Nowadays, the perception of users of VR applications and haptics is only measured by means of subjective evaluation [20]. These are time-consuming (human testers need to be involved), and expensive to conduct. Thus, even if these are necessary as a benchmark, there is a need for novel objective evaluation models, that can provide a measurement of QoE in real-time. In traditional 2D video streaming services (both on demand and real-time), this validation is performed by means of objective QoE metrics. Such are the cases of the state-of-the-art models for HTTP video streaming based on quality delivered, stalls, freezes and the end-to-end delay, for example [9]. Other examples include the frame-to-frame quality degradation assessments such as the Structural Similarities (SSIM) or the Video Quality Model (VQM) [20]. These objective models provide a grasp of how the video streaming services are degraded due to the effects of network and compression impairments. Enabling both the HAS and the frame-based objective methods to the VR video requires to take into account its multi-view nature, the error in the viewport prediction and the content. Some early approaches have appeared, such as [21]. However, there is still room for improvement and research. In addition, the objective quality metrics of the Tactile Internet need to be valid not only for the VR video feed, but also need to include the haptic communications. This is a widely unaddressed topic in the literature. Machine learning methods have shown promising results to predict QoE for 2D and 3D video streaming services [22]. These methods could provide an interesting starting point of research for the Tactile Internet VR use case.

D. Subjective QoE benchmarking and generic testbeds

Evaluation and benchmarking is fundamental to validate the strengths and weaknesses of the approaches under scrutiny. Given the novel nature of both VR video streaming and the tactile sensors, there is a need for new evaluation infrastructures. We have identified two areas of research: (1) subjective evaluation benchmarking and; (2) generic evaluation methods and testbeds.

Subjective evaluations provide the best measure of the quantitative quality of the content. For the current video streaming services, subjective evaluations are performed according to the recommendation ITU-T Rec.P.910, which focuses on multimedia tests [20]. Users are presented with the content and are asked to rate its quality following 5-level discrete scale ranging from bad to excellent (*i.e.*, the Mean Opinion Score - MOS). This subjective evaluation methodology is fit for 2D video streaming services. However, it becomes incomplete for the VR video streaming and even more when tactile sensors would be involved. The reason for this it that 360° video uses different (higher) resolutions and bitrates. Moreover, freely changing view brings observers a multi-view mode and immersive experience, which can not be provided by ordinary videos with the single point of view [23].

Finally, there is a need for standardization of VR evaluation infrastructures. We envision the development of testing infrastructures such as the one presented in Figure 2, in which the different components could be added step-by-step. In addition, there is a need for simulations and emulations environments where the scalability and robustness of Tactile Internet solutions can be evaluated. These are currently not available.

V. Conclusions

In order to achieve the real-timeliness (maximum 5 ms end-to-end latency) and quality (5 Gbps) required by the Tactile Internet, VR video streaming needs to overcome a number of challenges. The aim of this paper has been to pinpoint open research areas and challenges to enable VR for the Tactile Internet. After a theoretical and experimental evaluation of the state-of-the-art VR video streaming solutions, we have identified four areas of open research challenges. First, transport and application protocols need to be optimized and improved. We believe the research in this area goes in the direction of adaptive streaming where tiles, viewport prediction and infrastructure need to be thoroughly revised. Second, edge infrastructure will be fundamental to migrate computationally intensive tasks from the lightweight VR clients and to intelligently prefetch content to reduce latency. Third, the multi sensor system (VR+haptics) need to be synchronized. Finally, it is necessary to create new evaluation infrastructures and novel methods to subjectively benchmark solutions. We believe this work provides the means to open new opportunities for research not only within the challenging VR arena, but also in the field of management and control of wireless networks.

References

[1] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled Tactile Internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, March 2016.

[2] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 82–89, April 2017.

[3] G. Robles-De-La-Torres, "Virtual Reality: Touch/Haptics," *Encyclopedia of Perception*, vol. 2, pp. 1036–1038, 2009.

[4] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward Interconnected Virtual Reality: Opportunities, Challenges, and Enablers," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 110–117, 2017.

[5] A. Koonen, C. Oh, K. Mekonnen, Z. Cao, and E. Tangdiongga, "Ultra-High Capacity Indoor Optical Wireless Communication Using 2D-Steered Pencil Beams," *Journal of Lightwave Technology*, vol. 34, 2016.

[6] G. Torfs, H. Li, S. Agneessens, J. Bauwelinck, L. Breyne, O. Caytan, W. Joseph, S. Lemey, H. Rogier, A. Thielens, D. V. Ginste, J. V. Kerrebrouck, G. Vermeeren, X. Yin, and P. Demeester, "Atto: Wireless networking at fiber speed," *Journal of Lightwave Technology*, vol. 36, no. 8, pp. 1468–1477, April 2018.

[7] A. S. Shafigh, S. Glisic, and B. Lorenzo, "Dynamic network slicing for flexible radio access in tactile internet," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–7.

[8] R. Li, A. Clemm, U. Chunduri, L. Dong, and K. Makhijani, "A new framework and protocol for future networking applications," in *Workshop on Networking for Emerging Applications and Technologies (NEAT18)*. Budapest, Hungary: ACM, 2018, pp. 21–26.

[9] S. Petrangeli, J. van der Hooft, T. Wauters, and F. De Turck, "Quality of Experience-Centric Management of Adaptive Video Streaming Services: Status and Challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, November 2017.

[10] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, "An http/2-based adaptive streaming framework for 360° virtual reality videos," in *ACM on Multimedia Conference (MM17)*. Mountain View, California, USA: ACM, 2017, pp. 306–314.

[11] L. D´Acunto, J. van den Berg, E. Thomas, and O. Niamut, "Using MPEG DASH SRD for Zoomable and Navigable Video," in *ACM Multimedia Systems Conference (MMSys16)*. Klagenfurt, Austria: ACM, 2016, pp. 34:1–34:4. [Online]. Available: http://doi.acm.org/10.1145/2910017.2910634

[12] A. TaghaviNasrabadi, A. Mahzari, J. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using layered video coding," in *IEEE Virtual Reality (VR)*, 2017, p. 347348.

[13] G. Szabó, S. Rácz, D. Bezzera, I. Nogueira, and D. Sadok, "Media QoE enhancement With QUIC," in *IEEE Conference on Computer Communications Workshops*, San Francisco, CA, USA, 2016.

[14] S. Loreto and S. Romano, "How Far Are We from WebRTC-1.0? An Update on Standards and a Look at Whats Next," *IEEE Communications Magazine*, vol. 55, no. 7, p. 200207, 2017.

[15] P. Le Callet, S. Moeller, A. Perkis, and eds., "Qualinet white paper on definitions of Quality of Experience," in *COST Action IC 1003*, Laussanne, Switzerland, 2012.

[16] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, "BBR: congestion-based congestion control," *ACM Communications*, vol. 60, no. 2, pp. 58–66, January 2017.

[17] IETF, "rfc8257: DCTCP," {https://tools.ietf.org/html/rfc8257}.

[18] S. Zhao, Z. Li, and D. Medhi, "Low delay MPEG DASH streaming over the WebRTC data channel," in *Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, WA, USA, 2016, pp. 1–6.

[19] M. Eid, J. Cha, and A. E. Saddik, "Admux: An Adaptive Multiplexer for Haptic-Audio-Visual Data Communication," *IEEE Transactions in Instrument Measurements*, vol. 60, no. 1, pp. 21–31, 2011.

[20] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, June 2011.

[21] R. I. T. da Costa Filho, M. Torres Vega, M. C. Luizelli, J. van der Hooft, S. Petrangeli, T. Wauters, F. De Turck, and L. P. Gaspary, "Predicting the performance of virtual reality video streaming in mobile networks," in *9th ACM Multimedia Systems Conference (MMSys18)*. Amsterdam, Netherlands: ACM, june 2018, pp. 270–283.

[22] M. Torres Vega, D. Mocanu, S. Stavrou, and A. Liotta, "Predictive No-Reference Assessment of Video Quality," *Signal Processing: Image Communication*, vol. 52, p. 2032, March 2017.

[23] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei, "Subjective and objective quality assessment of panoramic videos in virtual reality environments," in *Conference on Multimedia & Expo Workshops (ICMEW)*, July 2017, pp. 163–168.