The Benefits of Caching and Computing for Wireless Virtual Reality

Ejder Baştuğ[◇,⊗], Mehdi Bennis[†], Muriel Médard[◇], and Mérouane Debbah[⊗,○]
[◇]Research Laboratory of Electronics, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA
[⊗]Large Networks and Systems Group (LANEAS), CentraleSupélec,
Université Paris-Saclay, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France
[†]Centre for Wireless Communications, University of Oulu, Finland
[○]Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France
{ejder, medard}@mit.edu, bennis@ee.oulu.fi, merouane.debbah@centralesupelec.fr

I. EXTENDED ABSTRACT

Just recently, the concept of augmented and virtual reality (AR/VR) over wireless has taken the entire 5G ecosystem by storm spurring an unprecedented interest from both academia, industry and others. Yet, the success of an immersive VR experience hinges on solving a plethora of grand challenges cutting across multiple disciplines. This work underscores the importance of VR technology as a disruptive use case of 5G (and beyond) harnessing the latest development of storage/memory, fog/edge computing, computer vision, artificial intelligence and others. In particular, the main requirements of wireless interconnected VR are described followed by a selection of key enablers, then, research avenues and their underlying grand challenges are presented. Furthermore, we examine a case study with numerical results, under various storage, computing and network configurations. This work exposes the limitations of current networks and makes the case for more theory, and innovations to spearhead VR for the masses.

A. Toward Interconnected VR

The overarching goal of virtual reality is to generate a digital real-time experience which mimics the full resolution of human perception. This entails recreating every photon our eyes see, every small vibration our ears hear and other cognitive aspects (e.g., touch, smell, etc.). It is envisaged that virtual reality (VR) systems will undergo three different evolution stages as depicted in Fig. 1, starting with current VR systems, evolving towards interconnected virtual reality (IVR), and finally ending up with the ideal VR system.

5G network architectures are being designed to move the post-processing at the network edge so that processors at the edge and the client display devices (VR goggles, smart TVs, tablets and phones) carry out advanced image processing to stitch camera feeds into dramatic effects.

To elaborate the context of current networks, even with a dozen or more cameras capturing a scene, audiences today only see one view at a time. Hence, the bandwidth requirements would not suffice to provide an aggregate of all camera feeds. To remedy to this, dynamic caching [2] and multicasting may help alleviate the load, by delivering content to thousands from a single feed. In a similar vein with the path towards user equipment (UE) centricity VR will instead let audiences

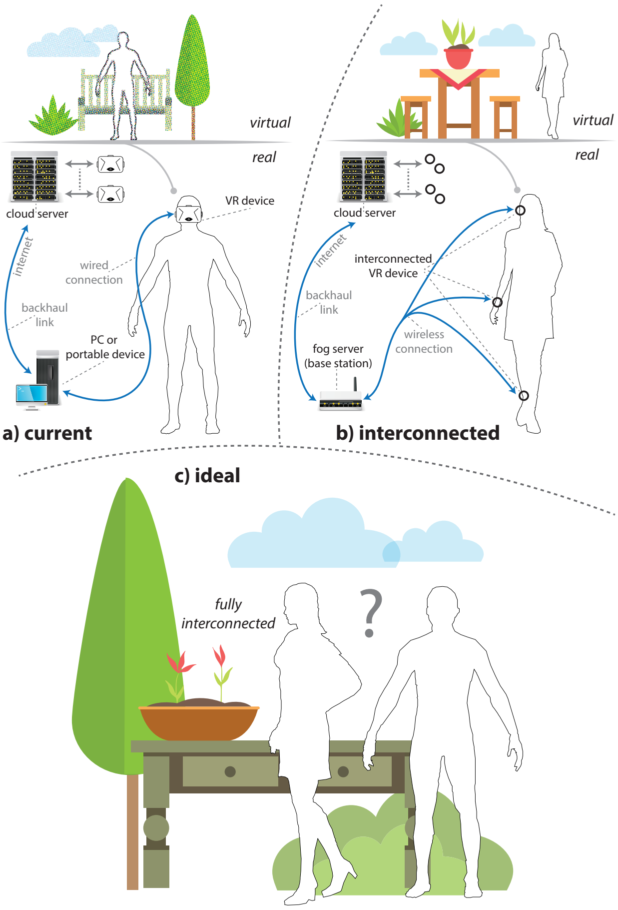

Figure 1: An illustration of virtual reality scenarios: a) current virtual reality systems, b) interconnected, and c) ideal (fully interconnected) systems.

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering), the U.S. National Science Foundation under Grant CCF-1409228, and the Academy of Finland CARMA project. This work is a short version of [1].

dynamically select their individual point of view. That means that the feed from *all* of the cameras needs to be available instantly and at the same time, meaning that conventional multicast will not be possible when each audience member selects an individualized viewpoint (unicast). This will cause outage and users' dissatisfaction.

In order to tackle these grand challenges, the 5G network architecture (radio access network (RAN), Edge and Core) will need to be much smarter than ever before by adaptively and dynamically making use of concepts such as software defined networking (SDN), network function virtualization (NFV) and network slicing, to mention a few facilitating a more flexible allocating resources (resource blocks (RBs), access point, storage, memory, computing, etc.) to meet these demands. Immersive technology will require massive improvements in terms of **bandwidth**, **latency** and **reliability**. Current remote-reality prototype (MirrorSys) requires 100-to-200Mbps for a one-way immersive experience. While MirrorSys uses a single 8K, estimates about photo-realistic VR will require two 16K × 16K screens (one to each eye). Latency is the other big issue in addition to reliability. With an augmented reality headset, for example, real-life visual and auditory information has to be taken in through the camera and sent to the fog/cloud for processing, with digital information sent back to be precisely overlaid onto the real-world environment, and all this has to happen in less time than it takes for humans to start noticing lag (no more than 13ms [3]). Factoring in the much needed high reliability criteria on top of these bandwidth and delay requirements clearly indicates the need for interactions between several research disciplines.

B. *Key Research Avenues and Scientific Challenges*

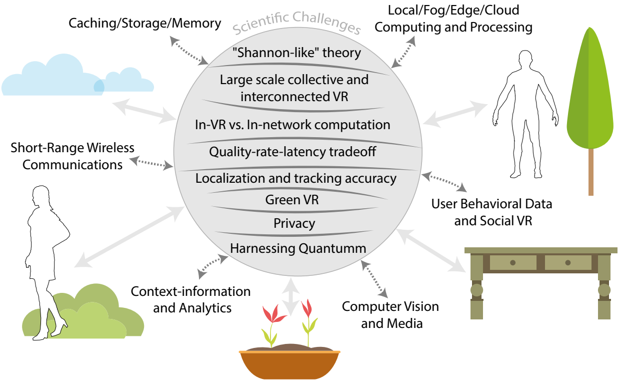

Figure 2: Research avenues and scientific challenges for interconnected VR.

The success of interconnected VR hinges on solving a number of research and scientific challenges across network and devices with heterogeneous capability of storage, computing, vision, communication and context-awareness. These key research directions and scientific challenges are summarized in Fig. 2. In addition, a numerical case study for AR-enabled self-driving vehicles under channel congestion is given in Fig. 3

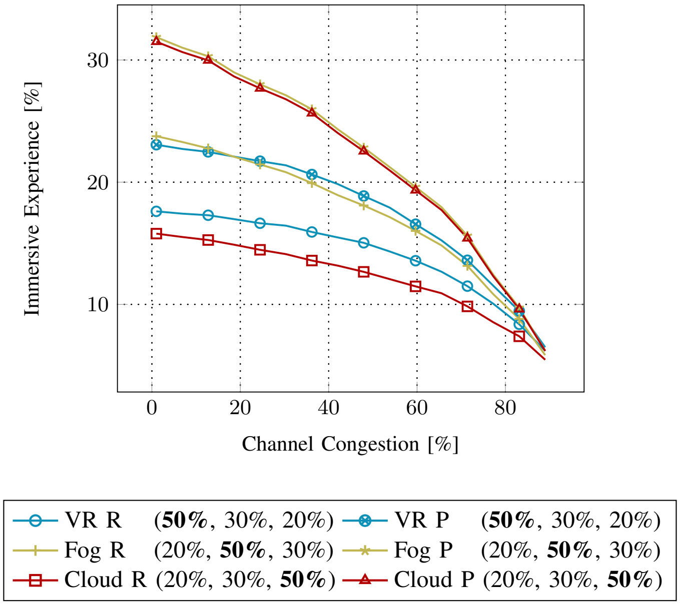

Figure 3: Evolution of the immersive experience with respect to the channel congestion, where fully reactive (R) and proactive (P) configurations of VR, Fog, Cloud-centric computation are considered. Fully reactive configuration has $S = 0\%$ of storage and the proactive configuration has $S = 80\%$ of storage. The default parameters are: Nr. of fog servers $M = 4$, Backhaul link capacity $L_{\text{ba}} = 64$ Mb/s, Wireless link capacity $L_{\text{wi}} = 1024$ Mb/s, VR computation capacity $C_{\text{vr}} = 4 \times 3.4$ GHz, Fog computation capacity $C_{\text{fg}} = 128 \times 4 \times 3.4$ GHz, Cloud computation capacity $C_{\text{cl}} = 1024 \times 4 \times 3.4$ GHz Simulation time $T_{\max} = 1$ hour, Average VR duration $T = 4$ ms, Average nr. of $\mu = 4$ tasks, Task arrival power-law exponent $\alpha = 0.8$, Average computation cost $\mu_{\text{co}} = 100$ Giga cycles, Computation cost power-law exponent $\alpha_{\text{co}} = 0.48$, Average delivery cost $\mu_{\text{de}} = 100$ MBit, Delivery cost power-law exponent $\alpha_{\text{de}} = 0.48$, Average deadline constraint $\mu_{\text{dl}} = 10$ ms, Deadline constraint power-law exponent $\alpha_{\text{dl}} = 0.48$.

C. *Ideal VR*

One speculative question which can be raised is whether an interconnected VR can reach to a maturity level so that no distinction between real and virtual worlds are made in human perception, making people to end up with the following question: *Are we living in a computer simulation?* In the context of VR, we call this unreachable phenomenon as *ideal (fully-interconnected) VR*. Indeed, we argue whether we can reach such a user experience with VR, therefore achieving an ideal (fully-interconnected) case. Despite the fact that we do now know the exact answer, we keep the ideal VR as a reference to all interconnected VR systems. We claim that the future lies in interconnected VR, despite its own research and scientific challenges.

REFERENCES

[1] E. Baştuğ, M. Bennis, M. Médard, and M. Debbah, "Towards interconnected virtual reality: Opportunities, challenges and enablers," *arXiv preprint arXiv: 1611.05356*, 2016.

[2] G. Paschos, E. Baştuğ, I. Land, G. Caire, and M. Debbah, "Wireless caching: Technical misconceptions and business barriers," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16–22, August 2016.

[3] M. C. Potter, B. Wyble, C. E. Hagmann, and E. S. McCourt, "Detecting meaning in RSVP at 13 ms per picture," *Attention, Perception, & Psychophysics*, vol. 76, no. 2, pp. 270–279, 2014.