2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 978-1-6654-8402-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/VRW55335.2022.00040

Reality Check of Metaverse: A First Look at Commercial Social Virtual Reality Platforms

Ruizhi Cheng* Nan Wu† Songqing Chen‡ Bo Han§

George Mason University

Abstract

Metaverse has grasped the news headlines recently. While being heavily advocated by the industry, there are great interests from academia as it demands various technological support from both hardware and software research. There has been an extensive visionary discussion of metaverse lately, but there are few studies on its technical challenges and user experience in practice. To fill this critical gap, in this paper, we take a first look at Workrooms, arguably a premature metaverse product released by Meta (*a.k.a.* Facebook). The goal of our study is to gain an in-depth understanding of the current state of the metaverse and identify potential issues for improvement. Through extensive measurement studies, we dissect the underlying network support and demand of Workrooms. Our investigation reveals several interesting findings. For example, Workrooms utilizes different network flows to transmit virtual content and real-time multimedia content separately. This might be a principled approach that should be adopted in general. On the other hand, the current design of Workrooms faces imperative scalability challenges that should be addressed in any large-scale metaverse platform.

1 Introduction

Metaverse, with the combination of the prefix "meta" (meaning transcending) and the word "universe", was coined by American writer Neal Stephenson in his 1992 sci-fi novel *Snow Crash*. Metaverse envisions a virtual reality-based successor to the Internet. In the novel, people use digital avatars to explore a shared virtual space that connects all virtual worlds via the Internet. Although the direction of building metaverse has gone through different stages after decades of development, social activities in the virtual world have always been considered the core element of the metaverse.

In recent years, with the flourishing of 5G and immersive computing [20], there has been a surge of research & development on metaverse in both industry and academia. Metaverse is considered to be a collection of 3D virtual worlds connected via the Internet and enabled by various emerging technologies such as blockchain, cryptocurrencies, and extended reality (XR), which includes augmented reality (AR), virtual reality (VR), and mixed reality (MR) [29]. Nevertheless, there is still no unified definition of the metaverse.

In this paper, we focus on one form of metaverse – social VR, the combination of online social networks and VR technologies. Social VR allows users to interact with each other as an avatar in the virtual world, communicating and collaborating as if they are in the physical world. It is regarded as the future of social media and an important component of the metaverse. At the same time, social VR has a salient practical value. With the global outbreak of the COVID-19 pandemic, many people around the world have to stay at home and lack social interactions, leading to a growing demand for novel applications of social media. Recent surveys show that 29.7% of respondents in the U.S. spend additional 1–2 hours a day on social media, while 51% of U.S. adults use social media at a higher rate during the COVID-19 pandemic [4, 8]. Predictably, the demand for social VR will continue to grow, as it not only satisfies people's social needs but also gives them a sense of spatial presence.

However, it is still not clear whether home networks can properly support social VR platforms. Currently, the U.S. Federal Communications Commission (FCC) defines the standard broadband service as 25 Mbps in downlink and 3 Mbps in uplink [31]. However, as can be expected, the bandwidth requirements of social VR platforms could be huge. On the one hand, compared to traditional 2D videos, the bandwidth requirements for transmitting 360-degree panorama or 3D volumetric content to AR/VR headsets are high [21, 28, 40, 48]. On the other hand, these platforms are full of social elements, which further increases the bandwidth requirements. Therefore, it is of the utmost importance to study the network requirements of existing social VR platforms to help us better understand the current state of metaverse development and build a more reliable network system for supporting metaverse.

In this paper, we conduct a first-of-its-kind measurement study of Workrooms, a newly released social VR platform from Meta (*a.k.a.* Facebook), which reflects the current industry efforts on metaverse. By dissecting Workrooms, we gain more insights into the current progress metaverse and identify its potential technical challenges. We conduct a series of experiments in a typical home network environment, by using several VR headsets (Oculus Quest 2) and PCs to study the social features of Workrooms. Our key findings are as follows:

- Workrooms primarily employs two servers to communicate with its clients, one using the UDP protocol for delivering virtual content and the other using WebRTC for streaming/exchanging audio and video data.
- With two VR headsets in Workrooms, each user's downlink throughput is about 2–3 Mbps and the uplink throughput is about 0.6 Mbps. However, the downlink throughput linearly increases with the number of headset users, indicating that the current design of Workrooms may face scalability issues.
- Workrooms does not consider situations that do not require server involvement (*e.g.,* peer-to-peer communication), but simply lets the server process and forward all users' data, resulting in unnecessary communication overhead.
- Each user receives two identical video sessions from the

*e-mail: rcheng4@gmu.edu
†e-mail: nwu5@gmu.edu
‡e-mail: sqchen@gmu.edu
§e-mail: bohan@gmu.edu

978-1-6654-8402-2/22/$31.00 ©2022 IEEE
DOI 10.1109/VRW55335.2022.00040

server, and the throughput and the number of sessions do not change when more users join Workrooms. However, we do not know exactly what is included in the video sessions and what is their data source. We hypothesize that it is related to real-time user activities (not confirmed yet).

The remainder of the paper is organized as follows. We present the background of the metaverse in §2, followed by an overview of several popular commercial social VR platforms in §3. We conduct a series of measurement studies in §4 to dissect Workrooms and understand its networking protocol usages and requirements. Finally, we review related work in §5 and make our concluding remarks, including the discussion of future work, in §6.

2 Background

In this section, we first briefly introduce the background of the metaverse and its current development in the industry. We then present the key challenges of building the metaverse from the network perspective.

Definition and Enabling Technologies. Metaverse has been viewed as a new type of online social network, or arguably the next-generation Internet. While there is no consensus on the definition, it is commonly agreed that metaverse is built on and integrates technologies such as 5G, immersive computing, edge computing, artificial intelligence (AI), and blockchain. Metaverse aims to provide users immersive experience based on AR, VR, and MR.

Objects in the physical world can interact with the metaverse by generating their digital twins through technologies such as 3D modeling. They can keep the digital twin presenting the same state as what is happening in the real world through sensors and other devices. Conversely, after the digital twin is manipulated/processed in the metaverse, its physical world state will be changed accordingly. For example, BMW, a world-renowned automobile company, has used the *Omniverse* platform to construct a fully functional, real-time automobile digital twin. It can simulate large-scale production and finite scheduling with constraints, reducing manufacturing costs and increasing productivity.

User-generated content (UGC) such as digital assets greatly enriches the metaverse. Defining the ownership of UGC in the metaverse is a practical challenge, as digital assets including digital art can be copied and reproduced. The non-fungible token (NFT) provides an effective way to prove the UGC is unique and non-fungible (*i.e.,* non-interchangeable) in the metaverse. NFT enables owners of digital content to sell/trade their property via smart contracts in the decentralized crypto space based on blockchain. As a concept introduced in 2014, NFT has been growing extremely fast in recent years. In 2021, its market value has reached more than $40 billion [3].

Current Industry Development. Many high-tech companies have joined the metaverse race. Meta is conceivably the most notable among all companies that have invested in this space. In September 2019, Meta (named Facebook then) announced *Facebook Horizon*, a VR social platform. In July 2021, it announced the transition into a metaverse company within five years. To echo this vision, in October 2021, it changed its name to Meta. Meta has invested $10+ billion to build the metaverse in 2021 and will continue the investment in the coming years. Meta considers VR as the foundation to build the metaverse. Meta's VR headset, Oculus Quest 2, has sold over 10 million units, making it the state-of-the-art and best-selling VR product in the world. Nvidia, on the other hand, announced a plan to create the first virtual collaboration and simulation platform called *Omniverse* in August 2021. This platform can be used to connect 3D worlds into a shared virtual universe and create digital twins, simulating real-world buildings and factories.

Platforms	Browser	Smartphone	PC App	Open Source
VRChat ('17)	✗	✗	✓	✗
Rec Room ('16)	✗	✓	✓	✗
AltspaceVR ('15)	✗	✗	✓	✗
Mozilla Hubs ('18)	✓	–	✗	✓
Workrooms ('21)	✓	✗	✗	✗

Table 1: Comparison of popular social VR platforms.

Although most companies embrace the metaverse's concepts and vision, cautions and doubts also emerge. While both Apple and Microsoft have virtual space applications[1], they consider that seamlessly connecting the metaverse and the physical world is a key to its success, if not more important than metaverse itself. They believe that the purpose of creating the virtual space is just to enable people to improve productivity and reduce production costs in the physical world. Thus, while some think the metaverse is the next-generation Internet, others believe the cyber-physical space (CPS) is more valuable than the metaverse. In their opinion, metaverse focuses on mainly the value of virtual space, while CPS emphasizes more on the value of bridging the virtual and physical worlds [2].

Technology Challenges. Although different companies have different views of the metaverse, it is undeniable that the metaverse is coming. Many agree that the metaverse is the next-generation Internet. Therefore, building a scalable, secure, reliable, and high quality of experience (QoE) network system is crucial to its success. In recent years, the boom in 5G has greatly enhanced the possibility of building reliable metaverse systems. 5G can reach a maximum throughput of, in theory, 10–20 Gbps. However, the bandwidth requirement of the metaverse is enormous due to the high-resolution video streams and huge metadata generated by the sensors. Considering the scalability demand, the bandwidth requirements of the metaverse may exceed what 5G can offer [29]. Meanwhile, network latency is crucial to the QoE. Low latency is particularly critical to motion sickness. Ensuring low latency when users are across geographically distributed regions is a practical challenge. Besides the high demand for network support, the security, accessibility, and economic aspects of the metaverse are also pivotal to the success of this new generation of the Internet.

3 Overview of Commercial Social VR Platforms

Since social VR is considered a major component of the metaverse, we provide an overview of several commercial social VR platforms in this section, highlighting their key features and differences. Admittedly, these platforms are in the different development stages towards a real metaverse.

Key Features. After an extensive survey, we focus on five of the most popular social VR platforms, VRChat [9], Rec Room [7], AltspaceVR [36], Mozilla Hubs [5], and Horizon Workrooms [35] (referred to as Workrooms). As a first step, we examine them from the following perspectives: i) Whether they are accessible from Web browsers? ii) Do they support smartphones? iii) Do they have PC applications? and iv) Are they open-source?

Table 1 presents a summary of these platforms. We find that most platforms currently have limited support for

[1] Apple acquired a VR company, Spaces, in 2020, and Microsoft acquired a social VR platform called AltspaceVR back in 2017.

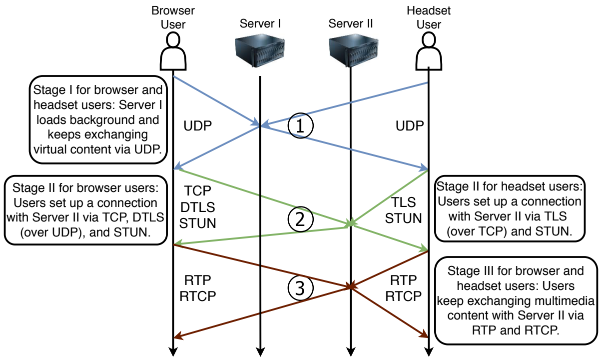

Figure 1: The process of establishing connections and exchanging data between the clients and the servers.

smartphones and browser-based user accesses. Only Mozilla Hubs and Workrooms offer browser-access options, while Rec Room is the only one to provide smartphone applications (Android and iOS). While Mozilla Hubs supports browser-based access on PC and mobile devices, Workrooms enables browser access on PC only. Furthermore, although most platforms have their PC applications for Windows, AltspaceVR is the only platform that has both Windows and macOS applications. Mozilla Hubs and Workrooms do not have PC applications, and thus users can use them through only browsers on PC. Finally, Mozilla Hubs is currently the only open-source social VR platform among them.

Among these platforms, Workrooms is the most recent effort on metaverse. In the next section, we will dive deeper into Workrooms. By exploring how Workrooms works, we can get a better understanding of the metaverse reality.

User Experience. We experiment with the above commercial social VR platforms and highlight their impressive advantages in terms of user experience.

- *VRChat:* Users can build their own games in the virtual world. It allows an impressive amount of customization (*e.g.,* users can upload any 3D model as the avatar).
- *Rec Room:* It enables cross-play between different users with different VR headsets, PCs, and smartphones. The interaction between users with different devices is smooth.
- *AltspaceVR:* The ambient lighting of the virtual scene matches the shadows, making the lighting of the scene very realistic. There are many environments and events initiated from all over the world, with a rich social element.
- *Mozilla Hubs:* Users can customize their own applications with its source code and deploy their own servers. They can use Hubs through browsers without downloading any application, which is lightweight and convenient.
- *Workrooms:* Users can use controllers by flipping them around and writing like a pen. It supports physical keyboards, which is much more convenient than the virtual ones manipulated by controllers.

4 Case Study: Dissecting Horizon Workrooms

In this section, we dissect how Workrooms works via a measurement study, mainly focusing on its underlying network protocols, highlighting our findings, and identifying areas to improve. As one of the key advocators of the metaverse, Meta (*a.k.a.* Facebook) released its product Horizon Workrooms [35] in 2021, which is a social VR-based remote collaboration platform that has many social elements in addition to the common features of video conferencing applications. For example, users can exchange ideas with each other on a virtual whiteboard, just like in a real meeting room.

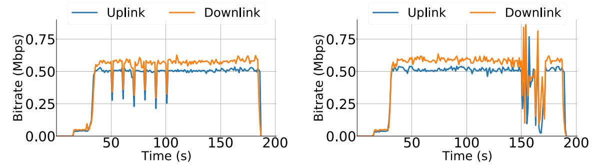

Figure 2: Bitrate of the UDP flow with the first server for U1 in G1 (the plots for U2 look similar). Left: switch background every 10 s from 50 to 100 s. Right: enable virtual whiteboard from 150 – 170 s.

4.1 Experiment Setup

We conduct a series of experiments with a 3-minute duration. We use a Macbook Pro as the WiFi access point (AP). It uses an external Ethernet adapter connected to the high-speed home network for Internet access. We capture and analyze the network traffic using the Wireshark packet analyzer [10]. We choose such an experimental setting in order not to cause performance degradation for three reasons. i) The downstream and upstream data rates of the WiFi AP are constantly greater than 70 Mbps and 50 Mbps, respectively. They are much higher than the downstream and upstream throughput required by Workrooms (§4.3, §4.4). ii) The network latency between our AP and the servers used for Workrooms is less than 10 ms (§4.2). iii) We conduct multiple experiments at different times of a day on different days over multiple weeks and report the most representative measurement result. Furthermore, we verify that Workrooms achieves the best performance using the OVR Metrics Tool, a performance monitoring tool of Oculus Quest 2 [6].

In order to investigate how different features and user actions, such as sharing PC screens and virtual whiteboards, affect the performance of Workrooms, we enable only one feature or perform one action in an experiment at a time and repetitively enable and disable that feature several times. Since Workrooms supports user access via either Oculus Quest 2 (the only VR device for Workrooms at this moment) or Web browsers, we conduct most experiments with two users under two settings. For the first setting (G1), both U1 and U2 use Oculus Quest 2 to access Workrooms, specified by G1-U1 and G1-U2, respectively. They use Quest 2 and Google Chrome, respectively, for the second setting (G2) and are designated as G2-U1 and G2-U2. Since a headset is required for creating a Workrooms team, we could not conduct experiments when both U1 and U2 use the browser.

4.2 Network Protocol Analysis

As the first step of our study, we aim to reverse-engineer the usage of network protocols employed by Workrooms. Figure 1 summarizes the process of establishing connections and exchanging data between the clients and the servers.

Through multiple experiments, we find that when the users are in the meeting room, their devices will communicate with two servers. The connection with Server I starts during the loading period (*i.e.,* when the loading progress bar is displayed). All data exchanges happen on UDP. We will show that this flow is for transmitting virtual content in the meeting room through a series of experiments in §4.3.

The connection with Server II starts when users enter the meeting room. The prefix of this server's hostname is "edge-mws-shv". The headset clients and the browser clients have a slightly different way of establishing connections with Server II. First, they both establish a TCP connection with Server II, while using Session Traversal Utilities for NAT (STUN) protocol to traverse network address translator (NAT) gateways. After that, the headset client and Server II transfer 1-3 Transport Layer Security (TLS) packets to each other. TLS is a secure communication protocol running over TCP. The

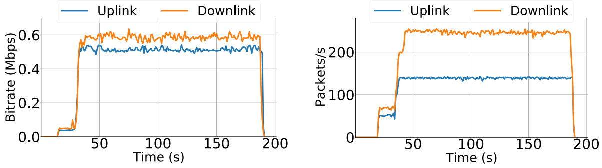

Figure 3: Bitrate (left) and PPS (right) of the VC flow for G1-U1.

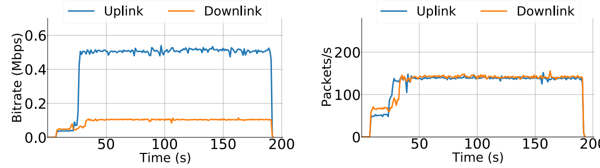

Figure 4: Bitrate (left) and PPS (right) of the VC flow for G2-U1.

TCP payload of these TLS packets is 110 bytes and 66 bytes for the downlink and uplink, respectively. However, after establishing a TCP connection with Server II, the browser client does not transmit any additional TCP packets, but establishes a Datagram Transport Layer Security (DTLS) connection with Server II. DTLS is a secure communication protocol over UDP.

After the connection is established, both browser and headset clients use Real-time Transport Protocol (RTP) and RTP Control Protocol (RTCP) to exchange multimedia content with Server II. RTP is used to transmit multimedia streaming (*i.e.,* audio and/or video), and RTCP is used to monitor data delivery. Through the certificates exchanged between the browser user and the server when they establish the DTLS connection and the Chrome WebRTC debugging console, we find that the server uses WebRTC (Web Real-Time Communication) [22] technology to transmit multimedia streaming for browser clients. WebRTC consists of a series of application programming interfaces (APIs) and communication protocols to enable real-time communication between the server and Web browsers (and/or mobile applications). It is used by many audio and video applications such as video conferencing and cloud gaming services [17, 31]. We speculate that the server also uses WebRTC for headset clients, although we do not have concrete evidence yet. In the G2 experiment, we find the synchronization source (SSRC) field in the RTP packets transmitted by the server to headset clients and browser clients are the same (§4.4), indicating that the video and audio streaming transmitted from the server to headset clients and browser clients are the same.

4.3 UDP Flows and Virtual Content Transmission

In this section, we investigate the UDP flow with the first server as described in §4.2.

To understand the purpose of this UDP flow, we experiment with two functions related to virtual content, change the background (*i.e.,* switching the virtual background of the meeting room) and enable a virtual whiteboard that allows users to write on it with their controller as a pen just like in a physical meeting room. Both functions require users to access Workrooms with a headset. Figure 2 (left) shows the throughput (*i.e.,* bitrate) of this UDP flow when switching the virtual background every 10 s from 50 to 100 s. During the switch, the throughput for both uplink and downlink of this UDP flow degrades and then quickly recovers. Correspondingly, we observe that each time when the background is switched, the headset's screen goes black briefly, and subsequently a new background is displayed.

Figure 2 (right) shows the throughput of this UDP flow when we enable the virtual whiteboard from 150 to 170 s. After the whiteboard is enabled, the throughput for both uplink and downlink no longer remains stable, fluctuating between 0.1 to 0.8 Mbps. This is due to the change of virtual content caused by the whiteboard. When the whiteboard is enabled, the headset's screen will have three parts: the whiteboard, the original meeting room, and the other user's avatar. These three parts are combined and rendered based on the user's current viewpoint, leading to the regeneration of the virtual background when the user moves and thus the fluctuation of the throughput. We also observe that when these two functions are enabled, they do not affect the WebRTC flows (§4.4). Through these experiments, we infer that this UDP flow is used to deliver virtual content. Thus, we refer it to virtual content (VC) flow hereafter.

Next, we study the throughput and packet rate (*i.e.,* packets per second, PPS) of VC flows and how they are affected by the number of users. We conduct the experiments when no function such as switch background or whiteboard is enabled. Figure 3 shows the bitrate (left) and packet rate (right) of these VC flows in G1. When the application starts, it takes about 20 s for loading where the throughput of uplink and downlink is below 0.1 Mbps. They then ramp up to 0.5 Mbps and 0.6 Mbps and remain stable till users exit the room after 3 minutes. The result of U2 is almost the same as U1 (not shown due to space limitations), implying that they send/receive the same (or the same type of) data.

Figure 4 shows the measurement result of VC flows in G2. By comparing Figure 4 with Figure 3, we observe that the uplink bitrate of G2-U1 does not change significantly, while the downlink bitrate drops from 0.6 Mbps to 0.1 Mbps. Similarly, the uplink PPS of G2-U1 almost does not change, while the downlink PPS drops from 250 packets/s to 140 packets/s. The reason is that we find, interestingly, the VC flow does not exist for G2-U2, the browser-based user. The biggest difference between G2-U2 and other users is that this user does not have an avatar[2] and cannot consume the virtual content with 6DoF (six degrees of freedom) motion (*i.e.,* the viewpoint of G2-U2 is fixed). We infer that the virtual content viewed by G2-U2 may be transmitted as a video session by the WebRTC flow (§4.4), which is the only downlink session for this user.

Since for G2-U1, there is no avatar of the other user in the meeting room, we hypothesize that the source of this 0.1 Mbps of data is the virtual background (*e.g.,* virtual seats in the meeting room, virtual building outside the meeting room, *etc.*). To confirm this, we conduct another experiment in the G1 setting, where only U1 joins the meeting room. Thus, G1-U1 can see only the virtual content, the same as G2-U1. We obtain the same results as in Figure 4.

Coincidentally, 0.1 Mbps equals the difference between the downlink and uplink bitrates in Figure 3. Therefore, we hypothesize that for G1-U1 and G1-U2, the downlink data should consist of two parts of the virtual content. One part is the virtual background generated by the server, which accounts for 0.1 Mbps, whereas the other part is avatar-related virtual content sent by another user, which accounts for 0.5 Mbps. To confirm this, we conduct another experiment. In addition to the original two users, we let three other headset users join in G1 at 50, 100, and 150 s, respectively. Figure 5 shows the measurement result of the VC flows for G1-U1. By comparing Figure 5 and Figure 3, we find that when a new headset user joins the room, the downlink bitrate of G1-U1 will increase by 0.5 Mbps, confirming our above conjecture.

[2]This user is simply visualized as a rectangle.

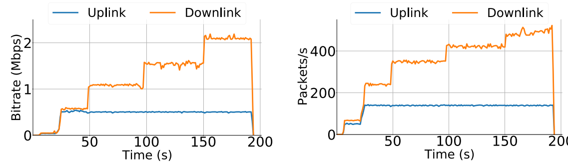

Figure 5: Bitrate (left) and packet rate (right) of the VC flow for G1-U1. U3, U4, and U5 join the meeting at 50, 100, and 150 s respectively.

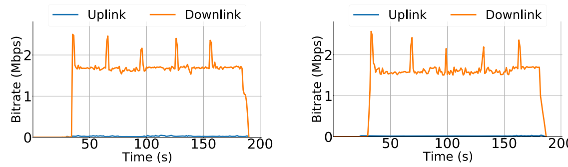

Figure 6: Bitrate of MM flows for G1-U1 (left) and G2-U1 (right).

Although the downlink PPS of G1-U1 will also increase with the addition of new headset users, its increase rate drops with more additional users. The reason is that when new headset users join the room, both PPS and packet size (not shown due to space limitations) will increase, leading to a drop of the increase rate for both PPS and packet size.

Key Findings: Through the analysis of this UDP flow, we have the following findings.

- This flow is mainly used to transmit the virtual content.
- Without activating any function, the virtual content in Workrooms contains two parts: the virtual background for about 0.1 Mbps data and users' avatar-related virtual content for about 0.5 Mbps data for each user.
- Workrooms simply forwards avatar-related virtual content to each user without further processing, which poses serious scalability issues. This is probably why it currently supports only up to 16 headset users. If there are 16 headset users, each user will receive ∼8 Mbps of data from the UDP flow. It is a significant bandwidth requirement and could be better designed and optimized.
- Workrooms does not take into account the case when the server does not need to forward data when the other user is browser-based, resulting in additional bandwidth overhead.

4.4 WebRTC Flows and Real-Time Streaming Content

Taking a similar approach as the analysis of the UDP flow, we verify that the WebRTC flows are for the delivery of audio and video content in real time. Such audio/video content is often related to real-time user activities. We refer to these flows as multimedia flows (MM for short) hereafter.

Bandwidth Requirement. We first explore the bandwidth requirement of the MM flows. Figure 6 shows the bitrate of the MM flows for G1-U1 (left) and G2-U1 (right), respectively. Unlike the VC flow, the MM flow transmits data almost exclusively on the downlink, with a bitrate of about 1-2 Mbps. The bitrate for the uplink is lower than 0.05 Mbps. We will verify next that the uplink data is mainly for audio sessions exchanged among users. In the same group, the downlink bitrate of U1 and U2 (not shown) remains almost identical, indicating they receive the same data. In some experiments, U1 and U2 receive almost the same amount of data from two different servers. The two servers are from the same region and have the same hostname and autonomous system number (ASN). We believe that this is due to the server load balancing policy. By combining the results in Figure 3 and Figure 6, we learn that the total bandwidth requirement for Workrooms is approximately 2 - 3 Mbps downlink and 0.6 Mbps uplink.

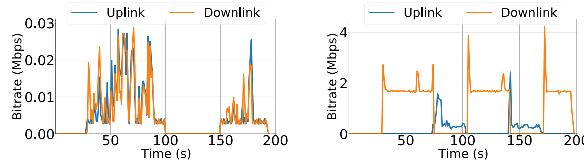

Figure 7: Bitrate of the audio sessions (left, both users mute from 100 to 150 s) and screen sharing (right, twice starting at 80 and 140 s, respectively) for G1-U1.

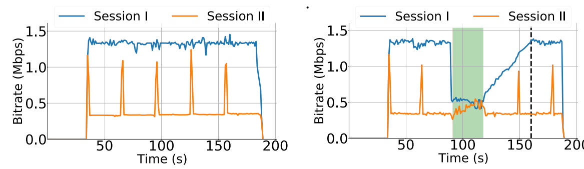

Figure 8: Bitrate of the two video sessions for G1-U1. Left: normal network condition. Right: limiting downlink throughput to 1 Mbps from 90 to 120 s (green region). The vertical dotted line indicates when the downlink returns to the original value before rate limiting.

Audio/Video Session Setup. Based on the SSRC field of RTP packets, we can distinguish between audio and video sessions. We find that Workrooms employs two audio sessions and two video sessions when there is no other function enabled. These two audio sessions are symmetric. Take the G1 experiment as an example. One audio session is sent from U1 to the server, and this session is subsequently forwarded by the server to U2. The other session is sent from the server to U1, and the source of this session is U2. Meanwhile, there are two video sessions, both of which are multiplexed by the server to U1 and U2 (*i.e.,* U1 and U2 receive the same two sessions). We observe the same in both G1 and G2. Furthermore, when the number of users increases, the number of video sessions does not change. Instead, each new user carries a new audio session, which is then forwarded to others by the server. This demonstrates that the video content in Workrooms is sent uniformly to all users by the server, while the audio data are relayed by the server to each user.

Audio Sessions. Figure 7 (left) shows the two audio sessions (*i.e.,* uplink and downlink) of G1-U1. To further verify that they are used for carrying audio data, both users mute themselves from 100 to 150 s, which leads to no traffic for these sessions during this period. The bitrate on the downlink of U1 exactly matches that of the uplink of U2 (not shown), and vice versa. Furthermore, when letting more users join the meeting, we find that the audio sessions between any two users present the same result. This indicates that the server simply forwards one user's audio data to others without further processing. However, given that the users are all in the same subnet, it is feasible to directly perform peer-to-peer (P2P) communication without the server forwarding. Workrooms currently does not employ P2P techniques to reduce the bandwidth overhead.

Video Sessions. After separating the audio sessions from the MM flows, we find, surprisingly, that there is no uplink data for video sessions. As mentioned above, increasing the number of users will not change the number of video sessions. Moreover, their throughput remains almost the same, when more users join Workrooms. Figure 8 shows the bitrate of the two sessions under normal network conditions (left) and with rate limiting (right). Under normal conditions, video session I maintains an almost constant bitrate, while the throughput of session II surges about every 30 s.

We hypothesize that session II may be used to probe the downlink capacity. To verify this, we limit the downlink bandwidth of G1-U1 to 1 Mbps from 90 to 120 s. As shown in Figure 8 (right), session II increases its bitrate during this period, which may indicate that it is indeed related to bandwidth probing. However, it does not help reduce the recovery time of video session I, which takes about 40 s to recover. For comparison, popular video conferencing applications such as Zoom and Google Meet take less than 20 s to restore downlink bitrate after disruption [31].

To summarize, it is not clear to us what exactly is included in the video sessions and what is the source of the video data. Our hypothesis is that it is related to real-time user activities, but this still needs to be confirmed. We are working on mechanisms to verify it.

Sharing Screen. We study how sharing the PC screen affects network transmission. This feature requires users to connect Quest 2 to a PC via the Oculus Remote Desktop application. Figure 7 (right) shows the measurement result. We find that when the user shares the PC screen at 80 and 140 s, a new video session is created to upload the screen data to the server, which then forwards the data to others. However, when a user starts to share the PC screen, the two original video sessions stop transmitting data. When users stop sharing the screen, the original video sessions will instantly transfer a large amount of data. When there are two users, this bitrate can reach about 4 Mbps (Figure 7). After repetitive experiments, we tend to believe that a lot of packet retransmissions occur at this time. The reason for this design is not clear to us yet, and we plan to investigate this issue in our future work.

Key Findings: Through the analysis of the MM flows, we have the following findings.

• The MM flows are divided into audio and video sessions. Audio sessions are forwarded between users by the server, while video sessions are sent to users by the server.

• Workrooms does not optimize the audio sessions via P2P communication, even when users are in the same subnet, resulting in additional bandwidth overhead.

• For video sessions that we infer to transmit real-time user activity related content (not confirmed yet), the server does adopt further processing before sending the data to the users, enabling it to support more users with better scalability.

• When sharing the PC screen, the original video sessions stop transmitting data and send a large amount of data instantly after the screen sharing ends.

5 Related Work

Metaverse. The metaverse concept has been discussed for about two decades [18, 23, 42, 46]. Recently, Jot *et al.* [25] proposed an efficient 6DoF spatial audio rendering solution for musical soundscapes. Duan *et al.* [19] presented a three-layer architecture and implemented a blockchain-driven metaverse prototype. Lee *et al.* [29] examined eight state-of-the-art technologies related to the metaverse and discussed its six user-centric factors. In this paper, we conduct a reality check of metaverse by measuring Workrooms, an early commercial prototype of the metaverse.

Social VR has recently attracted significant attention from the human-computer interaction (HCI) and VR communities, investigating issues such as locomotion and social mechanics [34], group interaction [38], personal space [14], augmentation of social behaviors [41], application design [24], prosocial interaction [33], avatar systems [26], non-verbal communication [32], *etc.* In the networking community, Zhang *et al.* [47] investigated the workflow of mobile social VR and proposed tentative system architecture. To the best of our knowledge, our study presents the first measurement study of commercial social VR platforms.

Mobile VR. There is a plethora of work on improving the performance and user experience of mobile VR [11, 15, 28, 30, 44]. Flashback [15] pre-computes and caches all possible frames on mobile devices, providing high-quality VR games. Furion [28] offloads costly background rendering to a server and performs only lightweight foreground rendering on mobile devices. MoVR [11] proposes a system that allows mm-Wave links to sustain high data rates even in the presence of a blockage and mobility. Different from the above work, in our study, we dissect the operations of a commercial social VR platform, Horizon Workrooms.

Online Social Networks. There is a rich literature on measuring various aspects of online social networks such as the structure of the underlying social group [37], user workloads/behavior [13], follower-following topological characteristics [27], user interactions [43, 45], and information propagation [16]. Besides the above measurement studies, Persona [12] is an online social network that protects user privacy through attribute-based encryption, which enables fine-grained policies over who may view what data. SPAR [39] makes online social networks scalable via a social partitioning and replication middleware. In contrast to the above work, we measure the next-generation online social networks, a social VR platform towards metaverse.

6 Concluding Remarks

While metaverse is up and rising in the media, much of the discussion has focused on the vision and its potential. Admittedly, metaverse could be built on top of the fast-developing 5G, AR/VR/MR, blockchain, HCI, and other technologies. The practical aspects of the metaverse, however, have received little attention. By studying a few commercial prototypes and particularly diving into Meta's Workrooms, we have conducted some preliminary experiments, aiming to gain a better understanding of the state-of-the-art and reveal potential challenges that may be faced in the future. Our measurement results demonstrate that scalability might be the most imperative issue to address.

As an initial exploration of the metaverse, our work has a few limitations. We have conducted measurement studies on only a single social VR platform, Workrooms. We plan to extend our efforts to other social VR platforms such as AltspaceVR and Mozilla Hubs and compare them with Workrooms. Workrooms currently supports only Oculus Quest 2, limiting the opportunity of investigating the impact of different VR devices on the system design and network performance. In this study, we have manually conducted the measurement studies. As it is in the early stage of the metaverse, there are few tools for automated testing of social VR platforms, especially using Oculus Quest 2. Although Oculus provides an automation testing tool called *AutoDriver* [1], it requires the implementation details of the applications. However, since the majority of social VR platforms are not open-source, we can perform only black-box measurement at this stage. We plan to develop open-source tools to enable large-scale, automated, and distributed experiments to help the future development of metaverse.

Acknowledgments

We appreciate the constructive comments from the reviewers. This work is supported in part by the NSF grant CNS-2007153, a Commonwealth Cyber Initiative grant, and 4-VA, a collaborative partnership for advancing the Commonwealth of Virginia.

References

[1] AutoDriver. https://developer.oculus.com/documentation/unity/ts-autodriver/, 2022. [accessed on 12-January-2022].

[2] Dialogue with Wei Qing, CTO of Microsoft China: What exactly is Metaverse? https://inf.news/en/tech/75c430c942f53df008547a695b88c5a2.html, 2022. [accessed on 12-January-2022].

[3] How NFTs became a $40bn market in 2021. https://www.ft.com/content/e95f5ac2-0476-41f4-abd4-8a99faa7737d, 2022. [accessed on 12-January-2022].

[4] Increased time spent on social by U.S. users during COVID-19 pandemic 2020. https://bit.ly/333uo75, 2022. [accessed on 12-January-2022].

[5] Mozilla Hubs. https://hubs.mozilla.com/, 2022. [accessed on 12-January-2022].

[6] OVR Metrics Tool. https://developer.oculus.com/documentation/unity/ts-ovrmetricstool/, 2022. [accessed on 12-January-2022].

[7] Rec Room. https://recroom.com/, 2022. [accessed on 12-January-2022].

[8] Uptick in US Adults' Social Media Usage Will Likely Normalize Post-Pandemic. https://bit.ly/3q8neXW, 2022. [accessed on 12-January-2022].

[9] VRchat. https://hello.vrchat.com/, 2022. [accessed on 12-January-2022].

[10] Wireshark. https://www.wireshark.org/, 2022. [accessed on 12-January-2022].

[11] O. Abari, D. Bharadia, A. Duffield, and D. Katabi. Enabling High-Quality Untethered Virtual Reality. In *Proceedings of USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2017.

[12] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: an online social network with user-defined privacy. In *Proceedings of ACM SIGCOMM Conference on Data Communication (SIGCOMM)*, 2009.

[13] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2009.

[14] A. Bönsch, S. Radke, H. Overath, L. M. Asché, J. Wendt, T. Vierjahn, U. Habel, and T. W. Kuhlen. Social VR: How personal space is affected by virtual agents' emotions. In *Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018.

[15] K. Boos, D. Chu, and E. Cuervo. FlashBack: Immersive Virtual Reality on Mobile Devices via Rendering Memoization. In *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services(MobiSys)*, 2016.

[16] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of International Conference on World Wide Web (WWW)*, 2009.

[17] A. Di Domenico, G. Perna, M. Trevisan, L. Vassio, and D. Giordano. A network analysis on cloud gaming: Stadia, GeForce Now and PSNow. *MDPI Network*, 1(3):247–260, 2021.

[18] J. D. N. Dionisio, W. G. B. III, and R. Gilbert. 3D Virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys*, 45(3):34:1–34:38, 2013.

[19] H. Duan, J. Li, S. Fan, Z. Lin, X. Wu, and W. Cai. Metaverse for Social Good: A University Campus Prototype. In *Proceedings of ACM International Conference on Multimedia (MM)*, 2021.

[20] B. Han. Mobile Immersive Computing: Research Challenges and the Road Ahead. *IEEE Communications Magazine*, 57(10):112–118, 2019.

[21] B. Han, Y. Liu, and F. Qian. ViVo: Visibility-Aware Mobile Volumetric Video Streaming. In *Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2020.

[22] C. Holmberg, S. Hakansson, and G. Eriksson. Web Real-Time Communication Use Cases and Requirements. RFC 7478, 2015.

[23] C. O. Jaynes, W. B. Seales, K. L. Calvert, Z. Fei, and J. N. Griffioen. The Metaverse: a networked collection of inexpensive, self-configuring, immersive environments. In *Proceedings of Eurographics Workshop on Virtual Environments (EGVE)*, 2003.

[24] M. Jonas, S. Said, D. Yu, C. Aiello, N. Furlo, and D. Zytko. Towards a taxonomy of social vr application design. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion (CHI PLAY), Extended Abstracts*, 2019.

[25] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt. Rendering Spatial Sound for Interoperable Experiences in the Audio Metaverse. In *Proceedings of International Conference on Immersive and 3D Audio (I3DA)*, 2021.

[26] A. Kolesnichenko, J. McVeigh-Schultz, and K. Isbister. Understanding emerging design practices for avatar systems in the commercial social vr ecology. In *Proceedings of ACM Conference Companion Publication on Designing Interactive Systems (DIS)*, 2019.

[27] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of International Conference on World Wide Web (WWW)*, 2010.

[28] Z. Lai, Y. C. Hu, Y. Cui, L. Sun, and N. Dai. Furion: Engineering high-quality immersive virtual reality on today's mobile devices. In *Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2017.

[29] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui. All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda. https://arxiv.org/abs/2110.05352, 2021.

[30] L. Liu, R. Zhong, W. Zhang, Y. Liu, J. Zhang, L. Zhang, and M. Gruteser. Cutting the Cord: Designing a High-quality Untethered VR System with Low Latency Remote Rendering. In *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services(MobiSys)*, 2018.

[31] K. MacMillan, T. Mangla, J. Saxon, and N. Feamster. Measuring the Performance and Network Utilization of Popular Video Conferencing Applications. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2021.

[32] D. Maloney, G. Freeman, and D. Y. Wohn. "Talking without a Voice" Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020.

[33] J. McVeigh-Schultz, A. Kolesnichenko, and K. Isbister. Shaping pro-social interaction in VR: an emerging design framework. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, 2019.

[34] J. McVeigh-Schultz, E. M. Segura, N. Merrill, and K. Isbister. What's It Mean to "Be Social" in VR?: Mapping the Social VR Design Ecology. In *Proceedings of ACM Conference Companion Publication on Designing Interactive Systems (DIS)*, 2018.

[35] Meta. Horizon Workrooms. https://www.oculus.com/workrooms/, 2022. [accessed on 12-January-2022].

[36] Microsoft. AltspaceVR. https://altvr.com/, 2022. [accessed on 12-January-2022].

[37] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2007.

[38] F. Moustafa and A. Steed. A longitudinal study of small group interaction in social virtual reality. In *Proceedings of*

ACM Symposium on Virtual Reality Software and Technology (VRST), 2018.

[39] J. M. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra, and P. Rodriguez. The little engine (s) that could: scaling online social networks. In *Proceedings of ACM SIGCOMM Conference on Data Communication (SIGCOMM)*, 2010.

[40] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan. Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices. In *Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2018.

[41] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018.

[42] B. Ryskeldiev, Y. Ochiai, M. Cohen, and J. Herder. Distributed Metaverse: Creating Decentralized Blockchain-based Model for Peer-to-peer Sharing of Virtual Spaces for Mixed Reality Applications. In *Proceedings of Augmented Human International Conference (AH)*, 2018.

[43] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2009.

[44] Z. Tan, Y. Li, Q. Li, Z. Zhang, Z. Li, and S. Lu. Supporting mobile VR in LTE networks: How close are we? *Proceedings of the ACM on Measurement and Analysis of Computing Systems (POMACS)*, 2(1):8:1–8:31, 2018.

[45] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *Proceedings of ACM European Conference on Computer Systems (EuroSys)*, 2009.

[46] M. Wright, H. Ekeus, R. Coyne, J. Stewart, P. Travlou, and R. Williams. Augmented duality: overlapping a metaverse with the real world. In *Proceedings of International Conference on Advances in Computer Entertainment Technology (ACE)*, 2008.

[47] L. Zhang, L. Sun, W. Wang, and J. Liu. Unlocking the door to mobile social VR: Architecture, experiments and challenges. *IEEE Network*, 32(1):160–165, 2017.

[48] W. Zhang, F. Qian, B. Han, and P. Hui. DeepVista: 16K Panoramic Cinema on Your Mobile Device. In *Proceedings of International Conference on World Wide Web (WWW)*, 2021.