2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW) | 979-8-3503-2812-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDCSW60045.2023.00025

Network Traffic in the Metaverse: The Case of Social VR

Ahmad ALHILAL *
Kirill SHATILOV *
Hong Kong University of Science and Technology
aalhilal;kshatilov@connect.ust.hk

Gareth TYSON
Hong Kong University of Science and Technology (Guangzhou)
gtyson@ust.hk

Tristan BRAUD
Hong Kong University of Science and Technology
braudt@ust.hk

Pan HUI
Hong Kong University of Science and Technology (Guangzhou)
Hong Kong University of Science and Technology
panhui@ust.hk

***Abstract*—The Metaverse connects our physical reality with virtual worlds. Social VR platforms facilitate the creation of such virtual worlds, enabling activities such as interactive teaching, conferences, and community gatherings. These activities can be performed in mixed mode, with some participants physically present in the same location. In this paper, we evaluate the feasibility of such mixed-mode events by studying three leading social VR platforms. We uncover the network traffic patterns generated by these platforms, which affect the user experience when multiple users share the same network. We explore the traffic patterns to show that model loading creates a significant overhead and impacts user quality of experience. When the number of simultaneously connected users increases, some operations lead to network congestion that degrades or even interrupts service for most users. From these observations, we derive concrete action points to develop social VR platforms that enable mixed-mode events over the same network.**

***Index Terms*—Virtual Reality, Metaverse, Social VR**

I. Introduction

As a growing range of social activities takes place in shared virtual environments, the border between cyberspace and the physical world blurs. Shared virtual environments, the building blocks of the Metaverse, enable concurrent interaction among multiple users with digital content, remote computing devices, and each other. In that regard, the Metaverse is an intersection of physical and virtual realities. Many social virtual reality (VR) platforms can create such virtual environments, such as Mozilla Hubs, Spatial and Meta Workrooms. These platforms create discrete "rooms", accessible on desktop, mobile, and VR head-mounted displays (HMDs). In these environments, users control 3D avatars that can move, talk, and interact with content. When using a VR headset, avatars also replicate the position and orientation of the users' major limbs. This increased user presence has led social VR platforms to be successfully used for running social activities, online conferences, and teaching courses [2]. However, most experiments so far have not accounted for mixed-mode operations where several users are accessing the virtual environment in the same physical location. With more events now happening in mixed-mode, we expect such a setting to become prevalent in the near future.

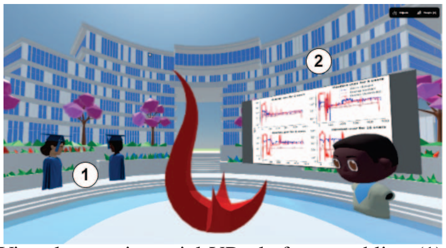
In July 2022, the Hong Kong University of Science and Technology organized one of the first mixed-mode press conferences in the Metaverse. During this press conference, the team set up a digital replica of the university campus in a Mozilla Hubs room that was accessed by over 30 reporters and presenters. With most invited participants having little experience in social VR, the event became an opportunity for a technological demonstration. The press conference took place in a single physical room, where guests were provided with VR headsets allowing them to join the virtual room. This article reports on our observations during a rehearsal performed one month before the press conference with 16 participants. During the rehearsal, we noted the significant impact of Mozilla Hubs' web-based architecture on the underlying network performance. Among other things, the large size of the 3D models led to significant delays to join and change scenes in the Hubs room, resulting in a poor user experience.

We set out in this article to explore the unique network traffic characteristics of Mozilla Hubs and compare its network footprint with Spatial and Workrooms[1]. After providing more context and motivation relative to the mixed-mode setting in Section II, we describe the measurement testbed we set up to instrument multiple virtual headsets and study the resulting network traffic (Section III). Building on this, we monitor the protocol messages to understand the stages of operation involved in running a Mozilla Hubs virtual environment. We show that network interaction in Hubs and Spatial consists of two key phases: (*i*) Model loading, which involves each device downloading the full 3D model; and (*ii*) Message exchange (via the server) between clients to synchronize interactions. We note significant costs in the first phase, particularly related to delays when moving between scenes. Therefore, we explore the per-device traffic exchange in Section IV. We show that, due to resource exhaustion at the server and access point, the model loading delay increases significantly as we introduce more clients. We then generalise our findings to

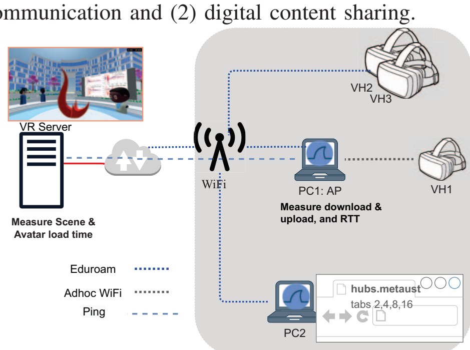
This work is supported in part by the MetaHKUST project from HKUST (Guangzhou).
The authors wish to thank Cheng et al. for providing data on their work [1].
*** A. ALHILAL** and **K. SHATILOV** contributed equally to this work.

[1] The data and analysis code are available on the following GitHub repo: https://github.com/ahmad-hl/NetTraffic_SocialVR

2332-5666/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDCSW60045.2023.00025



(a) Virtual scene in social VR platform enabling (1) VR communication and (2) digital content sharing.



(b) Measurement Testbed.

Fig. 1: Virtual room and measurement setup.

Spatial and Workroom in Section V. We expose that all three platforms suffer from the same issues that impede the quality of experience. We finally summarize our results and develop the concrete actions we took to host the press conference in satisfying conditions in Section VI.

II. Context and Motivation

This article explores the network characteristics of social VR platforms in the scope of mixed-mode events. After introducing Metaverse and social VR Platforms, we develop the scenario that led to our observations.

A. Background

The **Metaverse** is a term for technologies that blend physical and virtual realities. It brings together aspects of augmented, mixed and virtual reality to create shared virtual environments. There are many **social VR platforms** that support this concept, e.g. VRChat,[2] AltspaceVR,[3] Mozilla Hubs,[4]Horizon Workrooms,[5] and Spatial.[6] These platforms provide a lightweight 3D environment that can be rendered even on self-sustained VR headsets.

From a network perspective, only Meta Workrooms enables HTTP3, relying on QUIC over UDP.[7] Hubs, Workrooms and Spatial utilize WebRTC[8] for multimedia communication (enabling audio video streams) among the VR users. With the exclusion of Mozilla Hubs, these are all commercial platforms.

[2] https://hello.vrchat.com/
[3] https://altvr.com
[4] https://github.com/mozilla/hubs
[5] https://www.oculus.com/workrooms
[6] https://spatial.io/
[7] https://www.chromium.org/quic/
[8] https://webrtc.org/

B. Usage Scenario

Context In July 2022, the Hong Kong University of Science and Technology (HKUST) organised the first press conference in the metaverse[9] using Mozilla Hubs. The primary goal of this press conference was to communicate on HKUST's opening of its sister campus in Guangzhou, advertise HKUST's metaverse effort and introduce its metaverse campus to the public. The conference took place in a 3D reconstruction of the new campus, featuring its main piazza and the main research buildings. All the participants joined the Mozilla Hubs room using either VR headsets or laptops depending on their role. The press conference was conducted in mixed-mode, with more than 30 users present in the same physical location, connected to the same WiFi network.

Technical details This article reports on one of the first rehearsals that we conducted on 16 users. During this rehearsal, all participants were equipped with Oculus Quest 2 VR headsets connected to a single 2.4GHz WiFi access point on the eduroam network with a throughput of 78 Mb/s. We initially planned the press conference to happen in three consecutive phases, corresponding to three Mozilla Hubs scenes. The first phase was dedicated to learning how to use Mozilla Hubs, and took place in a mostly-empty area, featuring only 2 waypoints and one media frame to display instructional slides. As a result, this scene was very lightweight, less than 10 MB. After this tutorial phase, the Mozilla Hubs room would change scene to the model of the HKUST campus. This larger-scale scene (126 MB) featured the exterior of the main buildings of the new campus, with light baked into textures, and a few animations. Finally, the participants were moved to a final scene (5.4 MB), taking place in an amphitheatre where the presentations and speeches were to be held. Each participant was represented with a unique avatar of 2 to 5 MB in size.

Challenges During the first rehearsal, the participants encountered a major inconvenience. In Mozilla Hubs, all participants have to transition between scenes at the same time. The web browser also does not cache previously loaded 3D objects. Each scene change thus requires re-downloading all avatars along with the entire scene. Switching from the first tutorial scene to the second large-scale scene resulted in all 16 participants simultaneously downloading over 150 MB (100 MB map + 16 $\times$ 3 MB avatars) of data via the same access point. In optimal conditions, such a load would result in an average download time of 3 minutes and 12 seconds per user. Therefore, the participants' field of view in VR was replaced with a loading screen for an extended duration. As a result, participants started removing their headsets, triggering the headset's sleep mode. Upon entering sleep mode, the VR headset interrupts existing web connections, forcing the browser to re-download the entire scene and avatars. By removing and putting back their headsets, participants created a feedback loop that put additional strain on the network, preventing the rehearsal from progressing any further.

[9] https://hkust.edu.hk/news/research-and-innovation/hkust-launch-worlds-first-twin-campuses-metaverse

C. Contribution

While there has been much interest by institutions to transition to the Metaverse, it is yet to materialize at scale. As our experience shows, this is in part due to strict network requirements that make truly immersive environments difficult to support. VR remains a niche technology among a general audience, and mixed mode operation is likely to become the first introduction to the metaverse for many. However, such mixed mode events would only succeed if one can minimize the loading times for the users.

In this article, we explore the traffic profile of Mozilla Hubs to better understand the events leading to the network collapse described above. We specifically focus on the traffic between users and the Hubs server and compare it with network profiles of Spatial and Workrooms to highlight the common framework between them. We finally describe the concrete actions we took to prevent the issues that arose during the press conference and formulate recommendations for the deployment of future social VR and metaverse platforms.

III. Measurement Setup

The challenges observed within the rehearsal motivated us to systematically study the network footprint of VR platforms. To achieve this, we rely on a simple test environment, presented in Fig. 1b.

Virtual Room: We create a virtual room, as depicted in Fig. 1a. This is the main lobby of a university. We use this 3D room model across all experiments. This choice is inspired by our current work on building a virtual instantiation of a university in the Metaverse.

Server Setup: For Mozilla Hubs, the virtual room is hosted on AWS cloud in Virginia U.S. We select a C4 AWS EC2 instance[10] (c4.xlarge of 4 vcpu and 7.5 GB memory), which is optimized for compute-intensive workloads. The instance has a dedicated EBS bandwidth of 750 Mbps. The Hubs version is v1.1.4. For Spatial, we create a room by uploading the same 3D room model through its web interface.

Client Setup: Our testbed uses a Meta Quest 2 device (VH1), which we connect via a dedicated computer (PC1) to serve as the access point. We aim to characterize the traffic flow as users join and leave the VR room. Therefore, we include two further Meta Quest 2 client headsets (VH2, VH3). These are directly connected to a shared university WiFi access point. We choose a shared access point to study the traffic characteristics of co-located devices (e.g. several students in the same lecture theatre). Through these three devices, we investigate the upload and download traffic into VH1 as VH2 and VH3 join/leave the room. We use a WiFi network in a postgraduate lab on a Hong Kong university campus to access the Internet. The network to the VR server features an average bandwidth and round-trip time of 77.4 Mb/s and 220 ms, respectively[11]. PC1, serving as the access point, runs Wireshark to capture all traffic and records the per-second ingress and egress traffic between VH1 and the server.

[10] https://aws.amazon.com/ec2/instance-types/ [11] We use Iperf3 and ping to measure the bandwidth and round-trip time.

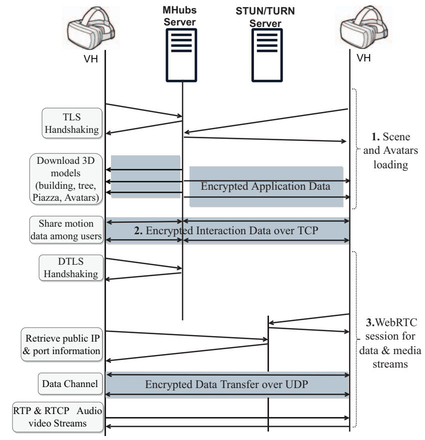


Fig. 2: Connection establishment and data exchange.

Emulating Load: To emulate concurrent access by multiple users, we introduce a second computer (PC2). We instrument this using Selenium to launch multiple tabs (2, 4, 8, 16) simultaneously. Each tab joins the virtual room to emulate the load on both the access point and AWS server. We also run Wireshark on PC2 to monitor traffic load.

IV. Case Study: Characterizing Hubs Traffic

We first focus on the network traffic of Mozilla Hubs, to understand the source of the issues that stemmed during the press conference rehearsal presented in Section II-B and mitigate their effect on the user experience.

A. Mozilla Hubs Network Operation Overview

We employ the above testbed to load our scene on headsets and monitor the packets between them and the server.

Connection Establishment Fig. 2 illustrates the observed flow of protocol messages between a headset and the Hubs server. Note, that all 3D rendering is done on the end-user device, while the Hubs server hosts the model and handles synchronization of positional data (virtual content), and the STUN/TURN server handles audio, video, and chat streams (real-time multimedia communication). The exchanges take place in four stages. The *first* stage establishes a secure (TLS) channel between the headset (client) and the Hubs server. This TLS connection is used to download the 3D models. *Second*, after model loading completion, the headsets exchanges encrypted interaction data (user pose and location) over TLS in the second stage. The *third* and final stage enables WebRTC for multimedia communication (audio and video) among the VR users through a STUN/TURN server. It beforehand establishes a secure channel for the data transfer over UDP (DTLS) among the users.

Multimedia Communication Protocols For the third stage (exchange of multimedia data), Hubs rely on direct peer-to-peer connections. This reduces the load on the Hubs server and also mitigates latency. Hubs uses WebRTC, a peer-to-peer protocol for real-time communications, for media streams

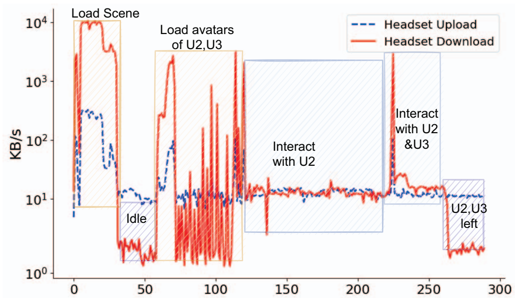
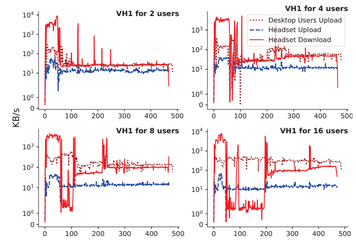
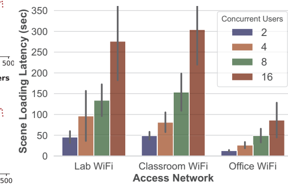


Fig. 3: Decomposition of traffic flow (download) of U1 as users (U2, U3) join the virtual room at 50 sec and 100 sec.



(a) Traffic rate of VH1 and sum of traffic rate of all clients.



(b) Scene loading latency over 5 runs over 3 WiFi networks.

Fig. 4: Impact of concurrent users on scalability.

between all users. Hubs uses a STUN server to establish connections and a TURN server to relay traffic if direct (peer-to-video gateway) connections fail with clients that do not have public IP addresses. Afterward, the clients exchange multimedia streams using the Real-time Transport Protocol (RTP), monitor the delivery, and adapt the bitrate using Real-time Transport Control Protocol (RTCP).

B. Traffic Flow Characterization

We first seek to characterize the traffic flows between the headset devices and the server. We rely on our testbed to instantiate a new Hubs room and have multiple devices connect. We run a simple experiment where we monitor all traffic between a single headset (U1) and the server. We then instrument two new users (U2 and U3) to join the room using two headsets (VH2 and VH3, depicted in Fig. 1b) after 50 seconds and 100 seconds, respectively. This is intended to understand how the sum download rate would change as further users join the room. We then measure the upload and download rates on PC1 for U1.

Fig. 3 presents the traffic rates (upload and download) from U1's headset (VH1) across the experiment. The figure clearly shows that the traffic varies to reflect the main tasks described in the prior section. Each of these is marked in Fig. 3. The first phase, *Model Loading*, takes 26 seconds to load in our experiment. This creates 137 MB of traffic with a traffic rate of 1$\pm$0.9 Mbps up and 37$\pm$32.2 Mbps. Note, we have a maximum available network capacity of 78 Mbps, as measured via Speedtest and the 3D model is 126.4 MB in size. This constitutes the peak traffic load across the entire experiment, with the majority of traffic being downloaded during this period. The U1's upload traffic increases slightly and is proportional (up $\approx$ 3% down) to the download rate because Hubs utilizes HTTP2/TCP for model loading, and sharing the users' interaction data. Since TCP is reliable and stream-oriented, acknowledgment and flow control packets contribute to the upload traffic besides the interaction data.

After loading completes, U1's download channel becomes idle (27^{th} - 60^{th} second), with an average traffic rate of just 0.08$\pm$0.03 Mbps up and 0.04$\pm$0.12 Mbps down. This is because, at this stage, there are no other users to interact with. Following this, we trigger *Avatar Loading*, when U2 and U3 join the room (at the 50^{th} second and 100^{th} second, respectively). This is because all clients must now download each others' avatars, resulting in a sudden peak of 23 Mbps, and a traffic flow of 0.15$\pm$0.15 Mbps up and 2.2$\pm$4.8 Mbps down. Thus, at this stage, U1 starts loading their avatars directly from the Mozilla Hubs server. Once this has been completed, it is possible for avatars to interact. Hence, we observe a relatively constant low traffic rate during the *Interaction Stage*, as clients exchange updates, e.g. avatar movement, and audio (0.104$\pm$0.011 Mbps up and 0.12$\pm$0.2 Mbps down). This takes place once U2 completes downloading the model on their device (125 – 225 seconds). Note, that interaction exchange requires substantially lower resources than the prior download phases with a relatively even balance between upload (sending updates) and download (receiving updates) traffic. As marked in the figure, we then manually trigger interaction between U1 with U2 and U3 (225 - 260 seconds). This results in a spike in download traffic (0.18$\pm$0.03 Mbps) as an update exchange takes place. Thus, we surmise that the download throughput grows linearly with the number of VR users, indicating potential scalability issues. Following this, U1's download rate returns to near idle when U2 and U3 leave the Hubs room (260 - 300 second).

Overall, although interactions do have relatively modest bandwidth requirements, low-capacity networks will suffer from substantial model loading delays. Besides, the download throughput increases linearly as a new user enters the interaction stage which may create scalability issues (e.g. a large number of users in low-capacity networks).

C. Impact of Concurrent Users

We next inspect the impact of multiple concurrent users on the above traffic profile. Our goal is to better understand how multiple participants might impact each other (e.g. in terms of avatar update exchange and network congestion). Using PC2, we introduce (2, 4, 8, 16) additional users to the Hubs room. We then measure the traffic between the VH1 headset (connected via PC1) and the server, as well as the traffic between PC2 and the server. Fig. 4a shows a time series of the traffic rates for VH1 across the experiments. These graphs contrast the traffic rate for the headset (red) vs. the rate for the desktop users on PC2 (dotted brown).

We derive several key observations from these results. We first see that the users are naturally competing for the available bandwidth via the shared access point. In Fig. 4a, we see that the duration of initial network traffic is extended as we increase the number of competing users. This occurs because our testbed access point has an upper bound of approximately 78 Mb/s. Thus, the initial delay of model and avatar loading grows as we increase the number of clients. For example, avatar loading takes just 3 seconds after scene loading for 2 competing users, but approximately 20 seconds for 4 users, and in excess of 100 and 200 seconds for 8 and 16 users.

A possible explanation is that each additional user forces all other participants to download an extra avatar. Fig. 4b plots the scene loading time over three access networks based on the number of concurrent clients (each configuration is repeated five times). The loading latency is primarily driven by the model loading time. Thus, it is largely agnostic to the number of avatars. Contention over network resources is the primary driving factor. We also notice some variance in the loading time across the five runs. This could be impacted by several factors, but we note that variations in signal strength degrade the results. With a signal strength of 55-70%, we observe a 50% drop in available bandwidth. This may create quality of experience issues when room joining times are synchronized (e.g. the beginning of a university lecture).

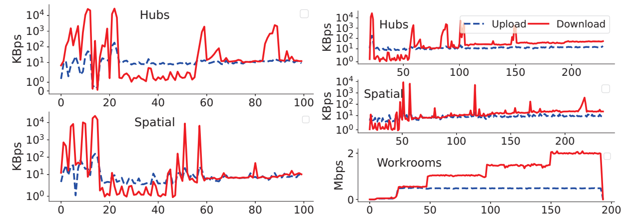
D. Per-Network Performance

The above experiments are all based on a single lab WiFi setup. To test how these generalize, we measure the scene loading latency in different access networks. Again, we run these experiments while testing different numbers of competing users (2, 4, 8, 16). We experiment with three separate WiFi networks: (*i*) WiFi in a shared office space; (*ii*) WiFi in a postgraduate lab on the university campus; and (*iii*) WiFi in a classroom on the university campus.

Fig. 4b presents the model loading latency for the three WiFi networks with different numbers of competing users (2,4,8,16 users). As expected, we see that the loading delay grows as we increase the number of competing users. Interestingly, this varies based on the network under study, with the office WiFi exhibiting the best performance. This is driven by the network capacity across these three sites. The office environment (181 ± 63 Mbps) has over twice that of the other two WiFi networks, Lab WiFi of average 76 ± 21 Mbps, and classroom with an average of 62 ± 22.8 Mbps. Unsurprisingly, this confirms that available access capacity has a major impact on loading times. To confirm that, we inspect the round-trip time for each access network. We use PC1 to ping the Hubs server and compute the network latency. RTT exhibits a low variance for the office WiFi, whereas it exhibits a high variance for the other access networks. This high variance indicates that the congestion resides in the access network.

E. Impact on User Perception

To investigate the impact of model downloading on user experience, we perform a user study with participants aged 19 to 36. We ask participants to use our Hubs setup before requesting a survey be completed. We obtain survey feedback from 12 users. Eight participants use VR a few times a year and three of them have never used it. The majority (10) are moderately or slightly familiar with VR.



(a) Rate on U1 for 2 users *(b)* Aggregated rate for 5 users.

Fig. 5: (a) Scene and avatar download patterns on U1 device. U2 joins the room at 50 sec. Spatial patterns are similar to Hubs'. (b) Average Aggregated download rate as U3, U4, U5 join at 50, 100, 150 sec, respectively.

Each participant wears a VR headset and joins the Hubs VR room (shown in Fig. 1a) over the Lab WiFi. We ask them to load the model and move around the environment, before a feedback survey about their perceived quality of experience. We use a simplified version of the NASA TLX [3] survey to obtain feedback, focusing on: (*i*) *frustration* on a [0-20] scale (0 being not frustrated at all and 20 being extremely frustrated); (*ii*) *mental demand* on [0-20] scale, (0 being easy and simple and 20 being demanding and complex). Besides the NASA TLX metrics, we also ask about the (*iii*) *perceived loading time* on [0-4] scale, (0: long and tedious, 1: somewhat long, 2: satisfactory, 3: acceptable, 4: fluent and fast). The participants reported low to medium frustration levels (average 7.25/20, sd 4) rates as low to medium mentally demanding (average 6.25/20, sd 4.4), and loading latency ranging from medium to high (average 3/4, sd 1). The perceived loading time ranges between long and acceptable, the setup appears to be relatively satisfactory but a notable fraction (4/12) found it excessive. Future work should therefore focus on optimizing this part of the experience.

V. Comparative Study of Social VR platforms

We generalize the previous results to two other commercial social VR platforms: *Spatial* and *Meta Workrooms*. These platforms operate similarly to Hubs, offering a browser-based VR experience with WebRTC for multimedia communication.

We upload the same Hubs 3D model (see Section III) to Spatial and set up a virtual room. We then use VH1 in our testbed (see Fig. 1b) to access the Spatial VR room. We rely on the work of Cheng et al [1] to characterize Workrooms. Only Meta Workrooms utilizes HTTP/3 for virtual data flow. Note, HTTP/3 significantly outperforms HTTP/2, especially in high latency or low bandwidth scenarios [4].

We compare traffic flows during the **interaction stage** and investigate how throughput scales with additional users. Fig. 5 illustrates the download throughput of the three selected platforms. We explore first the download patterns of scene and avatar through a browser for Spatial and Hubs. We monitor the downlink bitrate on U1's device as U2 joins the room at 50 sec, patterns are shown in Fig. 5a. We observe that Spatial exhibits scene and avatar loading patterns similar to Hubs. We then explore the downlink throughput as users U3, U4, and U5 join the room at 50, 100, and 150 seconds, respectively. The download rate increases slightly in Spatial (< 0.06 Mb/s), in Hubs (0.08 Mb/s), The download rate in Workrooms increases by the most (≈ 0.5 Mb/s).

A possible explanation is that the interaction data in Workrooms covers more avatar detail, e.g. (*i*) head motion, rotation, and pitch; (*ii*) facial expressions such as smiles, and eye motion; and (*iii*) hand gestures (not confirmed yet).

Thus, Workrooms have a marginally higher possibility of encountering scalability issues compared to Hubs. We observe that although Spatial captures and shares extra interaction data (e.g. simple facial expression, gesture, and body state) compared to Hubs, it is the most efficient in sharing interaction data (lowest downlink load). Spatial, therefore, can likely handle larger numbers of concurrent users in similar network conditions.

VI. Recommendations and Future Directions

Social VR platforms for education will face multiple challenges relating to the mode of deployment (all located in one space, mixed, fully decentralized), and user experience.

A. Main Observations

Metaverse users who are in the same physical location must compete for bandwidth at the shared access point. The model loading duration increases significantly in limited-capacity access networks, leading to longer waiting times. Besides, the downlink rate is the sum of virtual content traffic and real-time multimedia communication traffic (audio and video). This leads to degradation in real-timeliness when the rate exceeds the access network's capacity. This phenomenon is exacerbated by the increasing number of concurrent users in the same location. All three studied platforms exhibit the same behavior, with Spatial utilizing the bandwidth more efficiently (with a lower downlink bitrate), and Meta Workrooms requiring a download bitrate of 5-6$\times$ Hubs' download bitrate.

B. Deployment Recommendations

The three platforms suffer from the same long initial loading times that significantly affect user experience in VR. Although Spatial uses the bandwidth slightly more efficiently than Mozilla Hubs, we favoured the open-source component of Hubs in an educational context. As such, we took the following actions to ensure a seamless experience during the press conference:

Change network configuration. The first immediate action we took was to update the network. We upgraded the network throughput to 300Mb/s, disabled 2.4GHz operation, and connected all the headsets to the 5GHz band to minimize interference with other access points and users on campus. As a result, we noted a fourfold increase in the available bandwidth per user.

Reduce model size. Together with the increase in bandwidth, we decreased the model size in the second and third scenes. By favouring an entire low-polygon redesign, we managed to reduce the campus model from 126 MB to 45 MB and the amphitheatre model from 40 MB to 13 MB.

Adapt event organization. Despite the significant improvements in loading times, a 40 MB model with 30 unique avatars of average size 3MB loads at best in over a minute and a half on a 300Mb/s network for 30 concurrent users. As such, we made the decision to pre-load the heaviest scene on the headsets and let the participants train directly on this scene. This resulted in significant adjustments to the event organisation. First, as the VR headsets do not go to sleep, we had to provide power plugs at the location where each user would sit so that the headset keeps charging when not in use. Second, we had to perform individual training instead of group training due to the complexity of the scene. As such, we recruited 10 student helpers who were in charge of explaining the basic commands to a small number of participants and helping them if they encountered any difficulty.

C. Future Directions

As we observe that long waiting times lead to user frustration, future research should investigate the impact of the scene's 3D model size on user experience, and identify the trade-off between them such as optimizing the texture and reducing the size. They should also highlight the suitable settings for the different network conditions in users' environments (e.g. access from home or in school). Such research should assist to establish customized and fair Metaverse services.

VII. Conclusion

This paper investigated the network footprint of the Mozilla Hubs platform and compared it with two emerging VR platforms, Spatial and Meta Workrooms. We argue that these social VR platforms face a significant challenge related to avatars and model loading delays, exacerbated by multiple concurrent users in mixed-mode events. Limited access networks adversely affect the user experience, leading to unexpected side effects. Given these observations, we provided the concrete actions taken for preventing the network collapse during a real-life mixed-mode event, and advertised recommendations for the operation of future metaverse platforms.

References

[1] R. Cheng, N. Wu, S. Chen, and B. Han, "Reality check of metaverse: A first look at commercial social virtual reality platforms," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 141–148.

[2] L.-H. Lee *et al.*, "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," *arXiv:2110.05352*, 2021, unpublished.

[3] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[4] M. Trevisan, D. Giordano, I. Drago, and A. S. Khatouni, "Measuring http/3: Adoption and performance," in *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2021.