An In-depth Study of Bandwidth Allocation across Media Sources in Video Conferencing

Zejun Zhang
University of Southern California
Los Angeles, California, U.S.
zejunzha@usc.edu

Xiao Zhu*
Google
Mountain View, California, U.S.
shawnzhu@umich.edu

Anlan Zhang
University of Southern California
Los Angeles, California, U.S.
anlanzha@usc.edu

Feng Qian
University of Southern California
Los Angeles, California, U.S.
fengqian@usc.edu

Abstract

Video Conferencing Applications (VCAs) are indispensable for real-time communication in remote work and education by enabling simultaneous transmission of audio, video, and screen-sharing content. Despite their ubiquity, research on how these platforms allocate network bandwidth, especially under constrained conditions, and how these resource allocation strategies affect the users' Quality of Experience (QoE) is lacking. This paper addresses this gap by analyzing bandwidth allocation strategies in Zoom, Webex, and Google Meet, with a focus on QoE implications. To assess QoE, we propose a general QoE prediction model based on data collected from a study involving 800 participants. This study is a pioneering effort in evaluating multimedia transmissions across diverse scenarios and network conditions, advancing beyond prior research focused on single media types. The results demonstrate the model's effectiveness and generality in predicting QoE across various VCA scenarios.

CCS Concepts

• **Networks** → *Network measurement*; • **Information systems** → **Multimedia streaming**; • **Computing methodologies** → **Model development and analysis**.

Keywords

Quality of Experience; Multimedia Transmission; Video Conferencing; Measurement

ACM Reference Format:
Zejun Zhang, Xiao Zhu, Anlan Zhang, and Feng Qian. 2024. An In-depth Study of Bandwidth Allocation across Media Sources in Video Conferencing. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28-November 1, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3664647.3681007

*The work of Xiao Zhu was done when he was at the University of Michigan.

MM '24, October 28-November 1, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0686-8/24/10
https://doi.org/10.1145/3664647.3681007

1 Introduction

To enhance telepresence, VCAs have gradually integrated various media sources, including audio, video from camera streams, screen from screen-sharing streams, chat, and other advanced functionalities. This integration of multimedia transmission facilitates a highly customizable communication experience, enabling users to dynamically select and modify media inputs to suit their specific virtual meeting requirements. In real-world video conferencing, multiple media sources are often used simultaneously. For instance, teachers in online classes may use audio, video, and screen-sharing to provide a comprehensive learning experience.

Some prior studies analyzed the performance of popular VCAs [5, 16, 24, 29] and revealed their designs, including QoE metrics, network utilization, congestion control, etc. Others introduced innovative frameworks [4, 32] or systems [6] to enhance QoE. However, these studies mainly focused on individual media sources. A significant research gap remains in exploring bandwidth allocation across different media sources within VCAs, which is essential for optimizing performance and user satisfaction in video conferencing.

Under bandwidth constraints, video conferencing quality could degrade without careful resource allocation. The overall QoE depends on the combined performance of all concurrent multimedia sources. In scenarios with restricted bandwidth, how to allocate bandwidth—whether prioritized for one media source, divided equally among all media sources, or distributed unevenly—becomes critical in determining QoE. For example, in an online class with limited bandwidth, allocating all bandwidth to support screen-sharing clarity while ignoring audio transmission may make it difficult for students to follow the screen-sharing contents without clear audio, resulting in a reduced QoE. Instead, if each media source receives a proportionate share of the bandwidth to function at an acceptable level of quality, the overall QoE could be considerably enhanced. Therefore, investigating bandwidth allocation strategies that balance different media sources within network constraints is vital for optimizing the overall QoE.

To address this, we perform in-depth measurement and modeling of bandwidth allocation for three media sources: audio, video (camera streams), and screen (screen-sharing streams). Our study begins by examining the bandwidth allocation strategies of three major

commercial VCAs: Zoom, Webex, and Google Meet. Specifically, we focus on Zoom to examine its bitrate adaptation for each media source individually. Following this preliminary analysis, we conduct a broad user study to (1) explore the impact of different bandwidth allocation strategies on the QoE for real users and (2) develop a QoE prediction model general to various VCAs and scenarios. To the best of our knowledge, this model is the first to incorporate multiple media sources and serves as a benchmark to evaluate whether VCAs achieve optimal QoE in multimedia transmissions.

Navigating our research, we encounter several vital issues. First, acquiring QoE metrics like data rate, resolution, and latency from closed-source commercial VCAs is difficult. Second, to effectively gain real users' preferences from a user study, we need to design media source combinations that reflect a variety of network conditions. It is challenging to select a representative subset of these combinations for our user study while ensuring that we do not sacrifice the thoroughness and scope of our research. Third, building a general and robust QoE prediction model that applies to all scenarios and VCAs is essential.

Measurement of VCAs. To extract QoE metrics, we devise a measurement methodology to collect data from three VCAs. For large-scale controlled laboratory experiments, we engineered an automation tool responsible for client emulation, network control, and data aggregation at the client end. Among more than 20 hours of video sessions, we discovered their bandwidth allocation strategies and identified commonalities. Further, to explore the characteristics of individual media source transmission, we conduct an extensive case study on Zoom under restricted network conditions, specifically focusing on scenarios where bandwidth is limited and packet loss is high. Our significant findings are presented as follows:

• Under four scenarios with different combinations of media sources, the three VCAs consistently prioritize bandwidth allocation in the same order: Audio > Screen > Video.

• Zoom applies distinct bitrate adaptation strategies for video and screen. Specifically, it supports three-resolution video transmission and one-resolution screen transmission. However, this fixed strategy may not continuously satisfy user expectations under different scenarios.

User Study. Our IRB (Institutional Review Board)-approved user study successfully gathered 45,000 user ratings from 800 participants via Amazon Mechanical Turk [21] and covers four common usage scenarios. We formulate bitrate combination samples by merging different quality levels of three media sources.

Evaluating the QoE over such a wide range of bitrate combinations is challenging, primarily because comparing every possible combination with one another in a user study is impractical. To overcome this, we introduce an "accumulated score" method that allows us to compare two consecutive combinations as an alternative to comparing each possible pair. As a result, we can conduct a user study with only a fraction of the total combinations and still gain insight into user preferences across all possible pairs.

QoE Modeling. To interpret user ratings and define preference relationships, we employ the PageRank algorithm [2]. We then rank the combinations of media source bitrates based on the PageRank scores. Combinations that receive higher scores are identified as more preferred by users, establishing a clear preference hierarchy.

Following this, we develop a QoE prediction model that can be generalized to evaluate QoE across various VCAs and scenarios. This model is capable of predicting the QoE values for any given set of input combinations, enabling us to determine if an input combination achieves optimal QoE. Additionally, it allows us to rank a set of combinations, pinpointing which one offers the best QoE. This capability provides significant insights and actionable recommendations, guiding VCAs to improve user experience by fine-tuning their services to meet optimal user preferences, particularly in bandwidth-constrained environments.

Applying this model to evaluate Zoom, Webex, and Google Meet, we find that their performance is far away from the optimal QoE as predicted by our model. Among them, Zoom stands out by always offering a better QoE, showcasing its superior ability to manage bandwidth and adapt to varying network conditions. Nonetheless, all platforms have room for improvement to reach the optimal QoE.

The contributions of this paper can be summarized as:

• **Key observations and takeaways of VCAs.** We perform measurements of bandwidth allocation on three VCAs: Zoom, Webex, and Google Meet, providing valuable insights into their designs.

• **QoE Modeling.** We introduce a pioneering QoE prediction model that uniquely incorporates multiple media sources and is adaptable to various scenarios across different VCAs.

• **Measurement Tool and Crowd-sourcing.** We design an automated tool to programmatically control all experimental processes and conduct a large-scale user study to gather feedback from 800 participants.

This research does not raise any ethical issues.

2 Related Work

Measurement of Video Conferencing. Different Video Conferencing Applications (VCAs) use the same communication protocols but differ in their choice of codecs and traffic control strategies. This results in varied performance, even under identical network conditions. Macmillan et al. [16] measured Zoom, Google Meet, and Microsoft Teams, revealing distinctions in their recovery methods, video quality adaptation, and network utilization. Chang et al. [5] highlighted comparative results of streaming lag, audio/video QoE, and resource consumption among Zoom, Webex, and Google Meet. Yang et al. [29] evaluated system architecture, resilience to loss, and audio/video QoE for Google+, iChat, and Skype. Saini et al. [22] evaluated the performance of WebRTC-based Video Conferencing, including processing delay, CPU utilization, latency, jitter, packet loss, and packet delay.

QoE measurements are paramount when assessing VCAs. In terms of audio, commonly evaluated metrics include audio quality [16] and audio latency [29]. Video QoE assessments encompass aspects like framerate [15, 16], resolution [15], latency [29], and overall video quality [16]. Beyond these network-level analyses, researchers also delved deeper, employing transport-layer analysis to uncover the inner designs, such as congestion control [4, 12, 23], mechanisms for packet loss recovery [29], measurement-driven functional model [12], etc.

Some studies explored the security issues of VCAs. [8] conducted a dynamic security analysis of Zoom, Google Meet, and Microsoft Teams. [14] investigates three versions (desktop, web, smartphone) of Webex and identifies several relevant artifacts, including user

account information, encryption keys, media/text files, meeting records, etc. [18, 28] scrutinized Zoom's encryption method, offering insights and methodologies for decoding UDP and RTP packets.

QoE Modeling. [1] developed a predictive model of QoE for internet video. [7] conducted a small-scale user study to develop a QoE model for evaluating real-time video systems. [19] developed a QoE model to map network QoS metrics to video streaming QoE. [30] conducted a user study to model the QoE of 360-degree volumetric video streaming. However, these existing studies only focus on video sources without considering audio and screen-sharing media sources or their combined QoE.

3 Measurement of VCAs

In this section, we conduct a thorough analysis of three VCAs: Zoom, Webex, and Google Meet. Our focus centers on exploring their bandwidth allocation strategies for different media sources, including video, audio, and screen.

3.1 Measurement Methodology

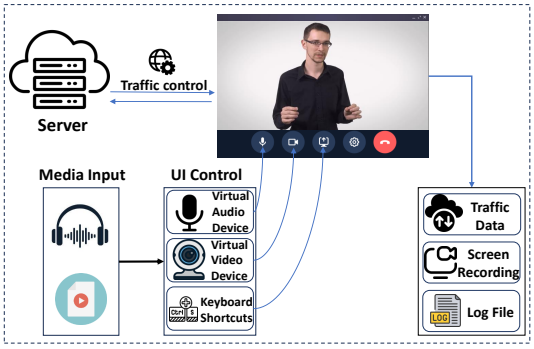


Figure 1: Testbed for measuring commercial VCAs.

An Automation Tool. To effectively control VCAs and simulate human activities programmatically, we develop a command-line automation tool, enabling efficient client emulation, network control, and data collection. It facilitates a streamlined process for conducting our experiments, as depicted in Figure 1.

• *Client Emulation.* To facilitate the automated sending and receiving of media sources, this tool incorporates *snd-aloop* modules and *aplay* [9] for audio input playback, along with *v4l2loopback* [10] modules paired with *FFmpeg* [26] for video input playback. We also utilize xdotool [25] to programmatically execute keyboard and mouse commands for various VCA operations, such as starting/ending screen, enabling/disabling audio/video, switching view layout, and opening/closing full-screen mode.

• *Network Control.* For managing network conditions on the client side, we use Linux *TC* [3], allowing us to configure uplink and downlink bandwidth and adjust latency precisely.

• *Data Collection.* For our analysis of bandwidth allocation across three VCAs, we capture network traffic via *tcpdump* [11] and obtain QoE metrics of each media source.

To collect QoE metrics for Zoom, we set up video or screen sharing in full-screen mode, with a statistics panel at the bottom-left corner displaying real-time resolution and framerate data, as shown in Figure 1. For the other two VCAs, which provide detailed log files, we periodically download and extract framerate and resolution averages. We preprocess input video/screen sources by overlaying a unique small QR code to each frame as the frame ID. We first recognize the QR code on each received frame to get the frame ID, then match frames with the same frame IDs between the sender and receiver. This process helps calculate the frame quality for each matched pair and determines the average video quality and framerate for the entire video conferencing session.

To obtain packet-level information, we decode packets on both the sender and receiver sides, extracting valuable details from the UDP/RTP headers. Since Zoom customizes its protocol, we refer to methods in [17, 18] to bypass the unknown datagrams.

Experimental Setup. Our measurement framework operates on machines running Ubuntu 22.04.1 LTS with Zoom 5.17.11, Webex 43.2, and Google Meet installed. These machines are connected to our on-campus wireless network, which guarantees a minimum bandwidth of 90 Mbps for both uploads and downloads. Each experiment includes $N (N \geq 2)$ users, where one user (referred to as "Sender") is responsible solely for uploading media to the VCA servers. Our research encompasses three measurements, each grounded in its unique experimental setup.

• *Bandwidth Allocation for Three VCAs:* In our study on bandwidth allocation across Zoom, Webex, and Google Meet, we identify four key scenarios reflecting different configurations and user behaviors, detailed in Table 1. These scenarios involve various media sources and window sizes. Our unidirectional experiments focus on evaluating the data rate of each media source under two conditions: limited uplink bandwidth at the sender and constrained downlink bandwidth at the receiver. Bandwidth limits are set at intervals of 0.2, 0.4, 0.6, 0.8, 1.0 Mbps. Each video conferencing session lasts five minutes, with each experiment conducted three times for reliability.

• *Zoom Measurement:* To examine the transmission behavior of each media source under network constraints, our experiments center on Zoom and involve one-directional tests with $N = 6$ participants. It includes one sender with unrestricted bandwidth, while the five receivers have unique downlink capacities, specifically *Unlimited*, *750Kbps*, *500Kbps*, *250Kbps*, and *150 Kbps*.

Media Inputs. For our audio input, we use a recording of a lecture where the lecturer engages in continuous speech, ensuring a consistent audio profile for the duration of our study. In terms of video and screen inputs, we select a lecture video, which is standardized to a resolution of 1280×720 and runs at a framerate of 25 FPS. To facilitate precise alignment of transmitted frames with their received counterparts, we embed a QR code for each frame of the video content. This methodological detail enhances the reliability of our frame-by-frame analysis.

	Audio	Video	Screen
Scenario 1	√	(Full-Screen)	
Scenario 2	√		(Full-Screen)
Scenario 3	√	(Thumbnail)	(Full-Screen)
Scenario 4	√	(Half-Screen)	(Half-Screen)

Table 1: Four scenarios with different media source inputs.

3.2 Network Utilization

VCAs employ distinct strategies for managing multimedia transmission. Before analyzing their performance under restricted network

	Audio (bps)		Video (bps)		Screen (bps)	
	send	receive	send	receive	send	receive
Zoom	100K	100K	1M	0.8M	1M	1M
Webex	95K	95K	700K	700K	1M	1M
Google Meet	70K	70K	1M	1M	800K	800K

Table 2: Data rate of different media sources in three VCAs.

conditions, we first measure their network utilization, focusing on the basic data rates of audio, video, and screen sharing in video conferencing. Table 2 displays the average data rates of these media types without any network constraints. For the three VCAs, the audio data rate is around 100Kbps, while the video and screen data rates are around 1Mbps. The sender's data rate is generally similar to the receiver's, except for Zoom's video transmission, where the receiver's data rate is 20% lower than the sender's. We explore this discrepancy further in §3.4.2.

3.3 Bandwidth Allocation across Media Sources

In practical video conferencing, using multiple media sources simultaneously is common. This section explores how VCA prioritizes and distributes bandwidth when multiple media sources are in play, especially under bandwidth-constrained conditions.

3.3.1 Scenario Descriptions.

We present four frequently encountered scenarios with distinct combinations of media sources, as outlined in Table 1.

• Scenario 1: Only audio and video connections are active, with the video in full-screen mode. An example of this scenario is an online interview, where body language and facial expressions are crucial.

• Scenario 2: Only audio and screen connections are active, with the screen in full-screen mode. This applies to group discussions, such as academic deliberations, where the focus is on slides or whiteboard content.

• Scenario 3: Audio, video, and screen connections are active, with the screen in full-screen mode and video displayed as a thumbnail in the upper right corner. This scenario commonly occurs in online conferences where a lecturer presents their work, and attendees primarily listen.

• Scenario 4: Audio, video, and screen connections are active, with the video and screen each occupying half of the window. A typical example is Big Tech companies' product launch events, where slides provide detailed information.

3.3.2 Zoom.

In Scenario 1, as bandwidth limits, the video data rate declines while the audio data rate remains consistent at around 100Kbps, as shown in Figure 2(a). At extremely low bandwidths (around 200Kbps), Zoom prioritizes audio, increasing its rate and nearly eliminating video. Scenario 2, illustrated in Figure 2(b), shows a similar trend but diverges as audio data rate increases when bandwidth narrows to 400Kbps, with screen data transmitting at a lower rate than audio even under severe constraints. Figure 2(c) shows that, in Scenario 3, tightening bandwidth stabilizes audio and video rates at approximately 120Kbps and 85Kbps. At 400Kbps, audio increases at the expense of other sources. At 200Kbps, video becomes negligible, with audio maintaining a higher rate than the screen. Scenario 4, as shown in Figure 2(d) shows both video and screen rates dropping as bandwidth restricts, while audio remains unchanged. At very low bandwidths, Scenario 4 trends similarly to Scenario 3.

3.3.3 Webex.

Figure 3 illustrates Webex's bandwidth allocation among three media sources. The audio data rate remains consistent, while video and screen-sharing rates decrease as downlink capacity is constrained. This mirrors Zoom's pattern, where under significant bandwidth limitations, the audio data rate surpasses those of screen-sharing and video, with video rates potentially approaching zero. This behavior suggests that Webex employs a similar traffic prioritization strategy to Zoom, favoring audio over other media types.

3.3.4 Google Meet.

Similar to Zoom and Webex, our analysis reveals that Google Meet prioritizes audio over video and screen-sharing in its traffic allocation, as shown in Figure 3. Additionally, in Scenario 4, the video data rate drops significantly even with moderate bandwidth (0.8Mbps). This suggests a deliberate strategy by Google Meet to maintain screen-sharing quality, potentially at the expense of video quality.

Takeaways. *Although three VCAs implement distinct bandwidth allocation strategies, they have the same bandwidth allocation prioritization: audio >screen >video. This fixed traffic prioritization for audio, video, and screen may degrade the user experience, as it may not match users' varying demands for these media sources based on their different meeting purposes.*

3.4 Case Study on Zoom

To understand individual media transmission, we conduct a case study on Zoom, focusing on examining Zoom's adaptive bitrate strategies in bandwidth-constrained networks.

3.4.1 Audio Transmission.

For audio-only conferencing, we observe that a consistent average bitrate of 100Kbps is maintained. We do not further apply bandwidth restrictions on audio transmission because we discover that a bandwidth lower than 150Kbps jeopardizes the stability of the meeting connection.

3.4.2 Video Transmission.

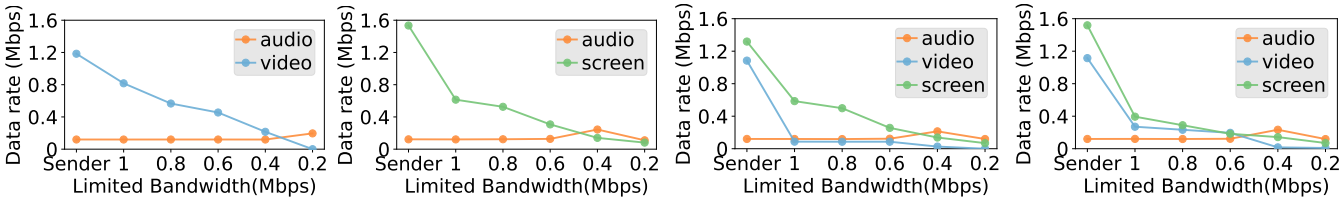
In multi-user video conferencing, declining downlink bandwidth at receivers leads to decreased data rates and varying QoE degradation. This degradation primarily affects framerate first, then resolution. Table 3 shows Receiver1, Receiver2, and Receiver3 maintaining resolution but with decreasing framerates as data rates drop. Further decline reduces resolution to 320×180 (180p) in Receiver4 and 256×144 (144p) in Receiver5, with a significant decline of VMAF and SSIM value, though framerates remain relatively stable.

	Sender (Unlimited)	Receiver1 (Unlimited)	Receiver2 (750Kbps)	Receiver3 (500Kbps)	Receiver4 (250Kbps)	Receiver5 (150Kbps)
Data rate (Kbps)	1158±120	883±130	647±85	453±44	218±35	144±20
Framerate (FPS)	21±3	21±3	13±2	10±2	8±1	7±2
Resolution	360p	360p	360p	360p	180p	144p
SSIM		0.89	0.87	0.84	0.82	0.8
VMAF		91	73	66	41	22

Table 3: QoE metrics of Video with bandwidth limits.

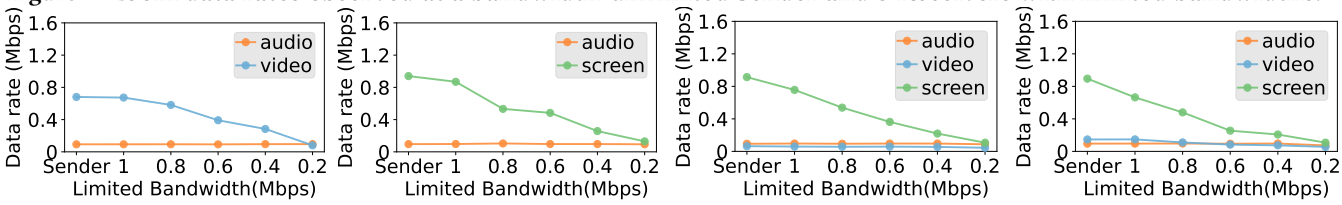
3.4.3 Screen Transmission.

Unlike video, the screen maintains a consistent resolution regardless of the downlink bandwidth allocated to each receiver. The resolution on the receiver side is the same as the sender side. If the sender's resolution changes, the receivers adjust accordingly. As bandwidth declines, the data rate and framerate drop correspondingly, as evidenced in Table 4. Intriguingly, even when the framerate



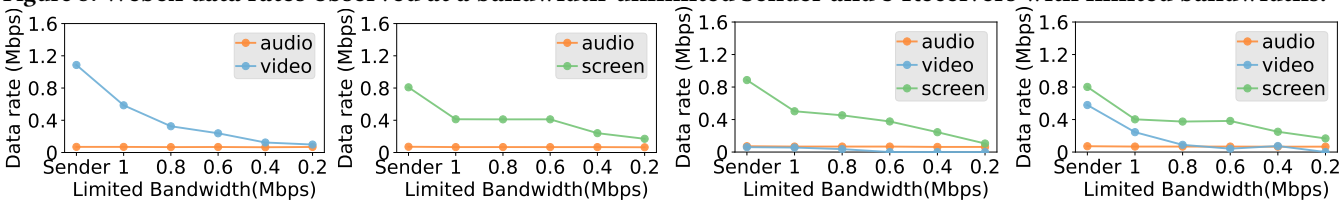
(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure 2: Zoom data rates observed at a bandwidth-unlimited Sender and 5 Receivers with limited bandwidths.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure 3: Webex data rates observed at a bandwidth-unlimited Sender and 5 Receivers with limited bandwidths.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure 4: Google Meet data rates observed at a bandwidth-unlimited Sender and 5 Receivers with limited bandwidths.

nears zero, the resolution remains unchanged across all receivers. Compared to Video, Screen-sharing's SSIM drops less as bandwidth decreases, indicating the clarity of each frame. This suggests that degradation in screen quality predominantly impacts the framerate.

	Sender (Unlimited)	Receiver1 (Unlimited)	Receiver2 (750Kbps)	Receiver3 (500Kbps)	Receiver4 (250Kbps)	Receiver5 (150Kbps)
Data rate (Kbps)	1482±230	1439±230	547±150	326±85	168±40	118±20
Framerate (FPS)	10±2	10±2	4±1	2±1	1±1	<1
Resolution	720p	720p	720p	720p	720p	720p
SSIM		0.91	0.89	0.88	0.87	0.85
VMAF		90	86	82	75	64

Table 4: QoE metrics of Screen with bandwidth limits.

Takeaways. *The three-resolution video transmission and one-resolution screen transmission don't adapt to various factors such as network conditions (*e.g.*, available bandwidth) and user configurations (*e.g.*, window size of the screen-sharing content), incurring network resource waste and QoE degradation. Thus, Zoom can strategically offload a part of the transcoding workload to more powerful Zoom servers, which also reduces uplink bandwidth usage, or find more intelligent adaptation strategies to balance the trade-off between the additional transcoding overhead and the quality/latency requirement.*

4 User Study

While VCAs have provided insights into their bandwidth allocation strategies under constrained network conditions, it remains unclear if these strategies align with user preferences or yield the optimal user experience. To bridge this gap, we begin with an IRB-approved user study to collect a dataset of real users' preferences on bandwidth allocation for diverse media sources in VCAs under constrained networks.

4.1 Methodology

Our user study methodology follows the Comparison Category Rating (CCR) method from ITU-T Rec. P.913 [20]. Instead of conducting an in-person user study, we opt for an online approach using Qualtrics [27] and Amazon Mechanical Turk (AMT) [21] to engage a globally diverse pool of participants. The study includes four scenarios, each with unique audio, video, and screen-sharing content. For each scenario, we generate video conferencing clips with different media quality levels, sorted by bitrates. Each clip lasts for 15 seconds.

Participants watch two side-by-side clips with consecutive bitrates. They can manually click the "Play" buttons to view each clip in full-screen mode and replay them multiple times before making a decision. Afterward, they subjectively compare their perceived QoE using a seven-choice scale ("The first one is {much better, better, slightly better, similar to, slightly worse, worse, much worse} than the second one."). For data processing purposes, these qualitative choices are converted into numerical values, with the scale translating to numbers from 1 to 7. To prevent audio interference between two video clips, participants are instructed to manually click the "play" button to view each clip sequentially.

To ensure rating reliability, we include several "test" comparisons between the best and worst clips. Data from participants who fail these "test" comparisons are discarded. Additionally, we track clicks and the time spent on each comparison, and remove data if the time is too short or there are fewer than three clicks.

Dataset Overview. To ensure broad applicability, our user study covers four representative video conferencing scenarios outlined

in Table 1, engaging 800 participants with demographics detailed in Table 5. Scenarios 1 and 2 involve 100 participants each, while Scenarios 3 and 4 have 300 participants each, yielding over 45,000 user ratings.

Age	18-25: 25.8%, 26-30: 27.0% 31-35: 16.4%, 35+: 30.8%
Gender	Male: 60.3%, Female: 39.2% Other: 0.5%
Country (30 Total)	US: 50.0%, IN: 30.1%, BR: 4.0%, IT: 5.7%, UK: 2.2%, Other: 6.1%
Education	Bachelor: 50.1%, Master: 26.3% Ph.D.: 8.1%, Other: 15.5%

Table 5: Demographics of the 800 subjects in our user studies.

4.2 Generating Bitrate Combination Samples

Given a specific bandwidth B, the potential bitrate combinations for distributing it among various media sources are infinite. Rather than attempting to enumerate an exhaustive list of these combinations, we need to strategically select a finite and representative set of bitrate combination samples. Inspired by Zoom's bitrate adaptation strategy, we create several quality levels for each media source. These differentiated quality levels of media sources are then combined, forming a selected set of bitrate combination samples.

We begin by producing benchmark media sources: an audio stream at 128Kbps, a video at 720p resolution with 25FPS and a bandwidth of 1.5Mbps, and a screen feed also at 720p and 25FPS consuming 1.5Mbps. Then, we transcode these benchmarks across a spectrum of quality levels. As shown in Table 6, we create 3 levels for audio and 9 levels (3 FPS levels × 3 resolution levels) for both video and screen. By combining different media sources together, we craft a set of 27 (3×9), 27 (3×9), 243 (3×9×9), and 243 (3×9×9) bitrate combination samples for Scenario 1, 2, 3, and 4, respectively.

Audio	128Kbps	32Kbps	8Kbps
Video	25FPS	15FPS	5FPS
	720p	360p	180p
Screen	25FPS	15FPS	5FPS
	720p	360p	180p

Table 6: Different quality levels of audio, video, and screen.

4.3 Calculating User Ratings

$$U = \begin{bmatrix} 0 & u_{1,2} & \cdots & \cdots & u_{1,n} \\ \vdots & 0 & u_{2,3} & \cdots & u_{2,n} \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & 0 & \cdots & 0 & u_{n-1,n} \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \quad (1)$$

To evaluate QoE across numerous bitrate combinations, we face the challenge of managing the vast number of pairwise comparisons. With N bitrate combination samples, the exhaustive pairwise comparison approach would necessitate $N(N-1)/2$ comparisons, which becomes unfeasible as N increases. Specifically, we need 2351 comparisons in Scenarios 1 and 2 and 29403 in Scenarios 3 and 4, which is clearly impractical. To address this, we propose the "accumulated score" method. It allows us to conduct N comparisons but still receives results that closely approximate results from $N(N-1)/2$ comparisons [13]. Essentially, this method enables us to deduce the entire user rating matrix U (as shown in Matrix 1) by examining only a fraction of its elements. It works as follows.

Ranking Combination Samples. We rank all N combination samples by their bitrate, based on the assumption that a higher bitrate typically means higher user preference. We compare each pair of adjacent combinations, N_i and N_{i+1}, where i ranges from 1 to $N-1$. This yielded $N-1$ user ratings, namely $u_{i,i+1}$.

Calculating Accumulated Scores. After comparison, we will get N user ratings ($u_{i,i+1}$). We set the accumulated score for the combination with the lowest bitrate (the N^{th} combination) to 0, namely $u_{N-1,N} = 0$. The accumulated score for the $(N-1)^{th}$ combination is calculated by adding the accumulated score of the N^{th} combination (acc_score_N) with the user rating obtained from the $(N-1)^{th}$ and N^{th} combination comparison ($u_{N-1,N}$). Accordingly, we apply this calculation sequentially to determine the accumulated scores for all N combinations by using the formula 2. These N accumulated scores are calculated on a per-user basis.

$$\begin{aligned} acc\_score_N &= u_{N-1,N} = 0 \\ acc\_score_i &= acc\_score_{i+1} + u_{i,i+1}, \quad i \in [1..N-1] \end{aligned} \quad (2)$$

Obtaining All User Ratings. After obtaining the accumulated scores for N combinations, we are able to determine the user rating between any two combinations by calculating the difference in their accumulated scores, as shown in Formula 3. This approach enables us to populate all the necessary elements in Matrix 1.

$$u_{i,j} = acc\_score_i - acc\_score_j, \quad i, j \in [1..N-1] \quad (3)$$

5 QoE Modeling

To understand user ratings and user preference relationships, we employ the PageRank algorithm [2] to establish a clear preference hierarchy and derive QoE values. Building on these insights, we create a QoE prediction model to predict QoE values for any given media source combinations.

5.1 QoE Values

The PageRank algorithm evaluates our user study results using a directed graph. Each node within this graph symbolizes a distinct bitrate combination, with edges between nodes representing comparative user ratings that highlight preference relationships. Here is a more detailed breakdown of the process.

• **Node Creation.** Each node in the graph corresponds to a unique media source bitrate combination. These combinations are directly derived from the scenarios presented in our user study.

• **Edge Construction and Weight Assignment.** The graph's edges are established based on the user ratings collected during the study. Participants are given seven options to express their preference between two combinations, *i.e.,* Combination A is *(1) much better*, *(2) better*, *(3) slightly better*, *(4) similar*, *(5) slightly worse*, *(6) worse*, *(7) much worse* than/to Combination B. These verbal options are then converted into a numerical scale of $\{3, 2, 1, 0, -1, -2, -3\}$, reflecting the degree of preference. An edge is drawn from node B to node A if the rating indicates a preference for Combination

A (rating > 0). Conversely, if the preference leans towards Combination B (rating < 0), an edge is drawn from node A to node B. The magnitude of the rating is used to assign weight to each edge, quantitatively expressing the degree of preference.

• **Assigning QoE Values.** PageRank [2] calculates scores for each node, effectively indicating the level of user preference for each bitrate combination sample. We then rank these combination samples, identifying those with higher scores as more favored by users. Following this ranking, we assign QoE values based on each combination's position in the preference hierarchy; the top-ranked or most preferred combination receives the highest possible QoE value, while the least favored combination is assigned a QoE value of 1. For scenarios 1 and 2, which feature 27 combinations, the QoE value for the highest-ranked combination is 27. Similarly, scenarios 3 and 4, each with 243 combinations, see their most preferred combination receiving a QoE value of 243.

5.2 Model Design

The input parameters for the QoE model, specifically designed to accommodate different scenarios, are detailed in Table 7.

Category	Parameter
Audio	[audio bitrate]
Video	[video resolution, video framerate]
Screen	[screen resolution, screen framerate]
Bandwidth	[overall bitrate]
Others	[ratio of window size between video and screen]

Table 7: Input parameters of each media source.

• Scenario 1: The input vector includes parameters specific to audio and video, along with the total bitrate.

• Scenario 2: This vector is associated with audio, screen-sharing, and the overall bitrate.

• Scenario 3 and 4: The input vector is all-encompassing, drawing parameters from every category, notably audio, video, screen-sharing, and the total bandwidth.

• General: The broad input vector aggregates parameters from all relevant categories—audio, video, screen-sharing, total bandwidth—and incorporates the newly introduced parameter of the window size ratio between video and screen-sharing. This approach ensures our QoE model's generality and relevance across different video conferencing scenarios.

In our analysis, we explore four distinct models, each meticulously adjusted to optimize our prediction task:

(1) Logistic Regression: We set $tol = 10^{-6}$, $random\_state = 0$, and $solver = newton\_cholesky$; This model is configured with a tolerance level of 10^{-6}, a $random\_state$ set to 0 for reproducibility, and utilizes the $newton\_cholesky$ method as its solver.

(2) Random Forest Regression (RF): We specify $n\_estimators = 100$, indicating the number of trees in the forest, and maintain a $random\_state$ of 0 to ensure consistent results across different runs.

(3) Gradient Boosting Decision Tree (GBDT): This model employs $n\_estimators = 10$, reflecting a more conservative approach with ten trees, and a $learning\_rate$ of 0.1, balancing the speed and accuracy of learning.

(4) Multi-layer Perceptron Regression (MLP): The MLP model is adjusted with a learning rate of 10^{-6}, employs a *logistic* activation function, an *adaptive* learning rate to adjust as learning progresses, a $random\_state$ of 0 for reproducibility, and $max\_iter = 2000$, allowing a generous number of iterations for convergence.

Our dataset is split into 80% training and 20% testing data. Using Python and the scikit-learn package, we employ four regression models and leverage ten-fold cross-validation (CV) [31] for training and evaluation. CV folds are determined at the user level to ensure unbiased results, as each user grades all videos and CV is performed separately for each user.

6 Evaluation

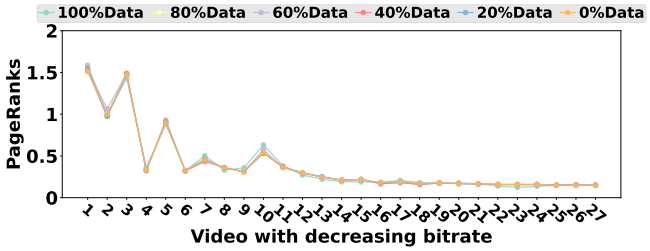


Figure 5: PageRank results with missing data.

6.1 Efficiency of Accumulated Score

This method evaluates consecutive combinations of media source bitrates instead of comparing every possible pair. We validate this approach through a simulation where five participants thoroughly assess each potential pair of combinations for Scenario 1, allowing us to create user rating matrices U directly.

For validation, we sequentially utilize 0%, 20%, 40%, 60%, 80%, and 100% of the user rating data in U, interpolating missing values using the "accumulated score" to create six corresponding matrices: U_0, U_{20}, ..., U_{100}. Then, we compute the PageRank for every combination sample across these six matrices, with results depicted in Figure 5. The similar trends observed indicate that using the "accumulated score" method aligns closely with the results of evaluating all possible pairs.

To evaluate the consistency of our method (U_0) with the approach of comparing every pair (U_{100}), we use the SequenceMatcher in the Python module to calculate the similarity in PageRank rankings between them. Achieving an average similarity score of 0.88 strongly affirms the effectiveness and reliability of our accumulated score methodology.

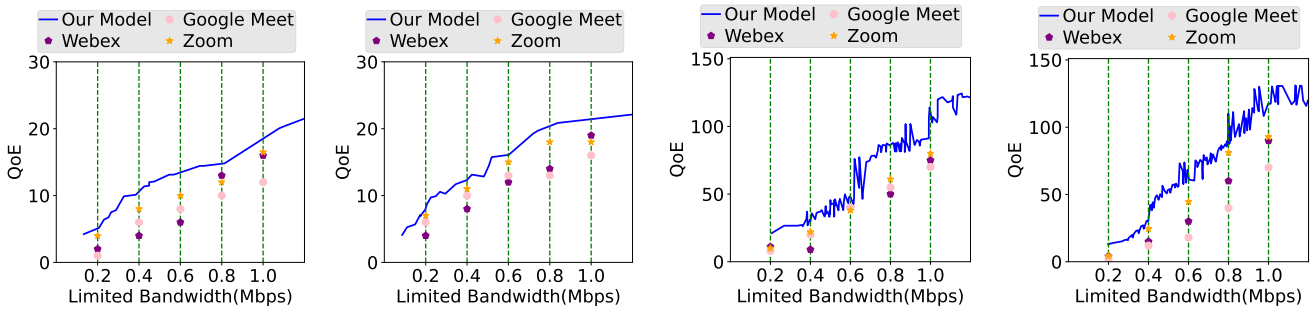
6.2 QoE Modeling Evaluation

Our QoE model is adept at predicting QoE values for specific combinations of media sources. This capability allows us to determine whether a given combination achieves the optimal QoE. Furthermore, when presented with multiple combinations, the model enables us to rank them based on their QoE performance.

• **QoE Prediction Evaluation.** We evaluate the accuracy of our QoE predictions using two key metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Lower values in these metrics indicate more accurate predictions, closely matching the actual QoE values from user studies. As shown in Table 8, all tested scenarios exhibit competitive MAE and RMSE scores, with Scenarios 1 and 2 demonstrating slightly better performance, likely due to the less complex nature of their bitrate combinations. Notably,

Scenario	Logistic			RF			GBDT			MLP		
	MAE	*RMSE*	*Accuracy*	*MAE*	*RMSE*	*Accuracy*	*MAE*	*RMSE*	*Accuracy*	*MAE*	*RMSE*	*Accuracy*
Scenario 1	0.12	0.92	82.13%	0.12	0.86	81.3%	0.15	0.93	82.1%	0.12	0.83	84.61%
Scenario 2	0.12	0.96	81.90%	0.13	0.87	82.15%	0.16	0.97	82.20%	0.11	0.85	84.55%
Scenario 3	0.15	2.81	78.79%	0.13	2.50	81.12%	0.16	2.87	81.04%	0.13	2.19	84.37%
Scenario 4	0.14	2.92	80.48%	0.11	2.28	82.06%	0.15	2.67	81.49%	0.12	2.24	82.62%
General	0.19	4.06	70.51%	0.09	1.75	81.63%	0.13	2.56	81.79%	0.08	1.78	82.86%

Table 8: Comparisons of the average MAE, RMSE, Accuracy (%) with Logistic, Random Forest, GBDT, MLP algorithms. The best results are underlined.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure 6: QoE comparison between Zoom, Webex, and Google Meet. Figures only show scenarios under 1.2Mbps.

the Multi-layer Perceptron (MLP) model achieves the lowest RMSE and MAE across all scenarios.

• **Combination Sets Ranking Evaluation.** We assess the model's accuracy in ranking various combinations by comparing its predicted QoE ranking to those derived from actual user feedback. According to Table 8, the models show promising performance overall, with accuracy exceeding 70%. The MLP model, in particular, stands out by consistently achieving accuracy rates above 80% in all scenarios.

In conclusion, our evaluation metrics underscore the model's effectiveness in precisely predicting QoE and in accurately ranking combinations. The MLP model surpasses other models in every scenario tested. Its robust performance across diverse scenarios highlights its generality and suitability for enhancing user experience within VCAs.

6.3 QoE Evaluation of Three VCAs

In §3.3, we investigate the bandwidth allocation of three VCAs: Zoom, Webex, and Google Meet. While this measurement provides valuable insights, it leaves the open question of whether these strategies truly align with user preferences or achieve the best possible user experience.

To address this, we apply our general QoE model (MLP model) to predict the QoE of Zoom, Webex, and Google Meet under restricted downlink bandwidth conditions (0.2, 0.4, 0.6, 0.8, and 1 Mbps). We use the QoE results from our user study as a benchmark (optimal QoE) to compare against the actual QoE performance of these VCAs.

As depicted in Figure 6, under the same bandwidth constraints, the performance of the three VCAs is far away from the optimal QoE calculated using the MLP model. Among these, Zoom demonstrates a higher QoE value compared to Webex and Google Meet in most scenarios, suggesting its bandwidth allocation strategies are more effective. Notably, the contrast between our benchmark and the predicted QoE values from the VCAs becomes more pronounced in Scenarios 2, 3, and 4 compared to Scenario 1. This significant difference is likely influenced by screen-sharing, which appears to affect the QoE outcome more severely in these scenarios.

7 Conclusion

Our research delves into the multimedia transmission capabilities of Video VCAs, with a particular focus on three key media sources: audio, video, and screen. Initially, we examine the bandwidth allocation strategies of three prominent VCAs—Zoom, Webex, and Google Meet—paying special attention to their performance in networks with limited bandwidth. Following this, we present a detailed case study on Zoom to explore its bitrate adaptation strategies for each media source when faced with network constraints about bandwidth limits and packet loss. Building on the above analysis, we propose a QoE model designed to predict QoE performance across various scenarios and platforms accurately. The findings from our evaluation demonstrate the model's effectiveness and generality. This model serves as a tool for VCAs to improve user experience by providing valuable insights and recommendations, particularly in scenarios with limited network resources.

However, our study still faces some limitations and requires further exploration. First, our QoE model focuses solely on bandwidth and does not consider other factors such as loss, jitter, and latency, which we plan to explore in future studies. Second, we currently use simple models, but we aim to design more complex and advanced models to achieve more accurate results as we incorporate additional factors. Third, the accuracy of prediction results is not very high because our model is designed to identify and cater to the most prevalent user preferences.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This research was in part supported by NSF under Grants 2409008, 2409267, and 2411625.

References

[1] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang. 2013. Developing a predictive model of quality of experience for internet video. *ACM SIGCOMM Computer Communication Review* 43, 4 (2013), 339–350.

[2] Pavel Berkhin. 2005. A survey on PageRank computing. *Internet mathematics* 2, 1 (2005), 73–120.

[3] bert hubert. [n. d.]. Linux TC Man Page. https://linux.die.net/man/8/tc

[4] Gaetano Carlucci, Luca De Cicco, Stefan Holmer, and Saverio Mascolo. 2016. Analysis and design of the google congestion control for web real-time communication (WebRTC). In *Proceedings of the 7th International Conference on Multimedia Systems*. 1–12.

[5] Hyunseok Chang, Matteo Varvello, Fang Hao, and Sarit Mukherjee. 2021. Can you see me now? A measurement study of Zoom, Webex, and Meet. In *Proceedings of the 21st ACM Internet Measurement Conference*. 216–228.

[6] Sandesh Dhawaskar Sathyanarayana, Kyunghan Lee, Dirk Grunwald, and Sangtae Ha. 2023. Converge: QoE-driven Multipath Video Conferencing over WebRTC. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 637–653.

[7] Sadjad Fouladi, John Emmons, Emre Orbay, Catherine Wu, Riad S Wahby, and Keith Winstein. 2018. Salsify:{Low-Latency} network video through tighter integration between a video codec and a transport protocol. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. 267–282.

[8] Nicholas Hunter Gauthier and Mohammad Iftekhar Husain. 2021. Dynamic security analysis of zoom, Google meet and Microsoft teams. In *Silicon Valley Cybersecurity Conference: First Conference, SVCC 2020, San Jose, CA, USA, December 17–19, 2020, Revised Selected Papers 1*. Springer, 3–24.

[9] Jens Getreu. 2017. *Redirect an audio stream with aloop.* Retrieved May 10, 2022 from https://blog.getreu.net/projects/snd-aloop-device/index.html

[10] Github. 2022. *v4l2loopback - a kernel module to create V4L2 loopback devices.* Retrieved May 10, 2022 from https://github.com/umlaeute/v4l2loopback

[11] Piyush Goyal and Anurag Goyal. 2017. Comparative study of two most popular packet sniffing tools-Tcpdump and Wireshark. In *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 77–81.

[12] Jia He, Mostafa Ammar, and Ellen Zegura. 2023. A Measurement-Derived Functional Model for the Interaction Between Congestion Control and QoE in Video Conferencing. In *International Conference on Passive and Active Network Measurement*. Springer, 129–159.

[13] Kevin G Jamieson and Robert Nowak. 2011. Active ranking using pairwise comparisons. *Advances in neural information processing systems* 24 (2011).

[14] Zainab Khalid, Farkhund Iqbal, Faouzi Kamoun, Liaqat Ali Khan, and Babar Shah. 2023. Forensic investigation of Cisco WebEx desktop client, web, and Android smartphone applications. *Annals of Telecommunications* 78, 3 (2023), 183–208.

[15] Insoo Lee, Jinsung Lee, Kyunghan Lee, Dirk Grunwald, and Sangtae Ha. 2021. Demystifying commercial video conferencing applications. In *Proceedings of the 29th ACM international conference on multimedia*. 3583–3591.

[16] Kyle MacMillan, Tarun Mangla, James Saxon, and Nick Feamster. 2021. Measuring the performance and network utilization of popular video conferencing applications. In *Proceedings of the 21st ACM Internet Measurement Conference*. 229–244.

[17] Bill Marczak and John Scott-Railton. 2020. Move Fast and Roll Your Own Crypto. https://citizenlab.ca/2020/04/move-fast-roll-your-own-crypto-a-quick-look-at-the-confidentiality-of-zoom-meetings/

[18] Oliver Michel, Satadal Sengupta, Hyojoon Kim, Ravi Netravali, and Jennifer Rexford. 2022. Enabling passive measurement of zoom performance in production networks. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 244–260.

[19] Ashkan Nikravesh, Qi Alfred Chen, Scott Haseley, Xiao Zhu, Geoffrey Challen, and Z Morley Mao. 2018. Qoe inference and improvement without end-host control. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 43–57.

[20] ITU-T P.913. 2021. Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.

[21] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making* 5, 5 (2010), 411–419.

[22] Shyam Sunder Saini and Lalit Sen Sharma. 2023. Performance Evaluation of WebRTC-Based Video Conferencing: A Comprehensive Analysis. *Journal of Advanced Zoology* 44 (2023).

[23] Constantin Sander, Ike Kunze, Klaus Wehrle, and Jan Rüth. 2021. Video conferencing and flow-rate fairness: a first look at Zoom and the impact of flow-queuing AQM. In *Passive and Active Measurement: 22nd International Conference, PAM 2021, Virtual Event, March 29–April 1, 2021, Proceedings 22*. Springer, 3–19.

[24] Ravinder Singh and Soumya Awasthi. 2020. Updated comparative analysis on video conferencing platforms-zoom, Google meet, Microsoft Teams, WebEx Teams and GoToMeetings. *EasyChair Preprint* 4026 (2020), 1–9.

[25] J Sissel. 2022. xdotool–fake keyboard/mouse input, window management, and more. https://github.com/jordansissel/xdotool

[26] Suramya Tomar. 2006. Converting video formats with FFmpeg. *Linux Journal* 2006, 146 (2006), 10.

[27] Sylvain Weber. 2021. A step-by-step procedure to implement discrete choice experiments in Qualtrics. *Social Science Computer Review* 39, 5 (2021), 903–921.

[28] Florian Wilkens, Steffen Haas, Johanna Amann, and Mathias Fischer. 2022. Passive, transparent, and selective TLS decryption for network security monitoring. In *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 87–105.

[29] Yang Xu, Chenguang Yu, Jingjiang Li, and Yong Liu. 2012. Video telephony for end-consumers: Measurement study of Google+, iChat, and Skype. In *Proceedings of the 2012 Internet Measurement Conference*. 371–384.

[30] Anlan Zhang, Chendong Wang, Bo Han, and Feng Qian. 2022. {YuZu}:{Neural-Enhanced} volumetric video streaming. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*. 137–154.

[31] Xinyu Zhang and Chu-An Liu. 2023. Model averaging prediction by K-fold cross-validation. *Journal of Econometrics* 235, 1 (2023), 280–301.

[32] Anfu Zhou, Huanhuan Zhang, Guangyuan Su, Leilei Wu, Ruoxuan Ma, Zhen Meng, Xinyu Zhang, Xiufeng Xie, Huadong Ma, and Xiaojiang Chen. 2019. Learning to coordinate video codec with transport protocol for mobile video telephony. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.