2020-03-10

An Evaluation of Zoom and Microsoft Teams Video Conferencing Software with Network Packet Loss and Latency

Alexander J. Clopper
Worcester Polytechnic Institute

Eric C. Baccei
Worcester Polytechnic Institute

Taylan J. Sel
Worcester Polytechnic Institute

An Evaluation of Zoom and Microsoft Teams Video Conferencing Software with Network Packet Loss and Latency

An Interactive Qualifying Project Report

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements

for the Degree of Bachelor of Science

Eric Baccei, Alexander Clopper, Taylan Sel

Submitted to: Professor Mark Claypool

Worcester Polytechnic Institute

Abstract

Video conferencing is an increasingly important form of communication, but there are few studies on how network quality affects new conferencing services. We studied Zoom and Microsoft Teams, video conferencing tools targeted towards businesses. Our user study had users rate their experiences during simulated normal conversation with network latencies and packet drop. Our study found that the network conditions did not significantly change user ratings, but Zoom had higher ratings than Microsoft Teams overall. Analysis of our benchmarking results showed that although network traffic for both services varied with network conditions, Zoom had a more constant bitrate than Teams which aligned with the responses received in the user study.

Table of Contents

Table of Figures

1 - Introduction

With the first experiments with video conferencing in the 1920s, it was considered a futuristic idea.[1] Video conferencing was in many science fiction forms of media, but seemed far from reality for the consumer. This all changed in the computer revolution of the 1980s with the increases in data communication infrastructure and the advent of video codecs. Additionally, the growth of mobile phones and webcams began to spur the growth of the video call and conference. One of the first widely available cameras, QuickCam, was even named as one of the most important technological advancements of the early 2000s. Although multiple versions of video calling services emerged over the years, Skype seemed to gain the most ground due to its free, cross platform service. Today, there are dozens of services for video calling, voice over IP communication, and video conferencing.

As a business, video conferencing has quickly grown. In 2018, it was estimated that revenue related to video conferencing was over 7.8 billion dollars (US), with some 32.8 million video conferencing devices.[2] Video conferencing is expected to grow up to 13.82 billion dollars by 2023. There are now multiple competitors for all markets. Both Microsoft Teams and Zoom are services for the business sector. These have the capacity to host a large number of users at once in the same room, the ability to share screens, and create leaders; all useful tools for running digital meetings.[3] However, the technology is being used for more than just meetings. Hospitals can use video calling and conferencing to interact with patients, even to help facilitate new ways to treat mental health.[4] Educators are also making full use of video calling. Even governments use these services now, with customers such as the Department of Homeland Security.[5]

Zoom is a video conferencing service designed for a range of businesses, with the ability to conference with up to 1000 people at once, and up to 49 people can display their screens at once.[3] Zoom also allows for versatility via connectivity options, including a mobile app, downloadable client and web client. Zoom has also expanded its features to include administrative options and a chat feature which make it ideal for business functions. Zoom has expanded its base to nearly 13 million active users as of February, 2020.[6]

Microsoft Teams is also designed for businesses, although more as an all in one tool rather than just for video conferencing. Teams offers a variety of tools, not all related to calling. The Teams platform can do file storage, file collaboration, and has a calendar application with reminders. Microsoft Teams also has many educational uses, such as being used to assign assignments and quizzes, and allows for text, video and audio from student to teacher.[7] Teams' audience has grown to 20 million active users as of November, 2019.[8]

Most studies on video conferencing are older studies with older technologies. For example, Skype has had extensive research done on it, both as voice over IP service and as a video calling service.[9][10] One such study was conducted by Xinggong Zhang and Yang Xu,[9] studied the effects of packets and bit rates on Skype, with user input for the quality of video. Batu Sat and Benjamin Wah compared Google Talk with Skype in terms of audio quality.[10]

Those previous studies are either out of date or do not focus on other technologies such as Zoom and Teams. In addition, many of the studies revolving around Skype, as well as other services, use an older version of the system, where connections were established peer to peer. This form of communication between computers has been largely replaced by client-server technologies. New technologies may alter how video call quality changes with changes to latency and packet loss.

Our methodology includes connecting three devices to the same network via Ethernet connection for two different studies. The three people in the call would then tell one joke each, and discuss the jokes. This was considered one “round”. Packet drop chance and latency levels were varied for each round using Clumsy. After each round, we asked the participants a set of questions to gather information on their experience with the quality of the call. For the benchmarking study, we used Wireshark to monitor incoming and outgoing packets during the connection. We used the same levels of packet drop and latency as with the user study.

We found that for all levels of latency and packet drop, Teams had consistently lower ratings than Zoom in terms of user experience. We saw that the user experience, when only latency was applied, did not change for either Teams or Zoom. The same was seen for Zoom and packet drop. However, when packet drop was applied to Teams, we did see a drop in user experience. For our benchmarking results, we saw that latency had little effect on kilobytes per second for either service.

Chapter 2 explains all of the background information necessary for this study. This includes some previous studies on services like Skype, in order to give a baseline for expectations. Chapter 3 explains our methodology and Chapter 4 lists our results from the study. Chapter 5 contains our conclusions and Chapter 6 explores future work that could be done.

2 - Background and Related Work

Relevant background for video conferencing research includes video compression, the different ways that clients are connected, how data is transferred, and tools which will be used for video conference analysis.

2.1 - Video Compression Technology

In order to efficiently transmit video data on smaller bandwidths, the frames that make up every electronic video must be encoded and compressed. Coding techniques in these compression algorithms allow for less redundancy in successive frames.[11] There are two primary types of video compression, the older MPEG and the newer H.261/263.[12]

2.1.1 - MPEG

The first video streaming technologies primarily used the MPEG structure. Under MPEG, the video feed is split into three different types of frames. Each frame is a compressed still image. For example, consider a video of a man raising his arm from a resting position to above his head. The I frame, or intra frame, is the frame that MPEG streaming is based upon.[12] Suppose a data stream for a video of a man slightly moving his arm. From I1 to I2, for example, the man's arm goes from at his side to above his head. These are two separate single frames, encoded as still images. In between these I frames are predicted frames and bidirectional frames, P and B frames respectively. Both “predict” by calculating the difference between the frames they refer to. The P frame tries to “predict” the movements in between the I frames by encoding the difference between these I frames. The same happens with the B frames as well, but here a distinct difference is made. The B frame encodes differences looking both forward and backward, at both the previous I and P frames and the future I and P frames. Figure 1 shows this “predictive” encoding in greater detail. We can see that in B frames are the most common, and that many P and B frames are used for every one I frame.

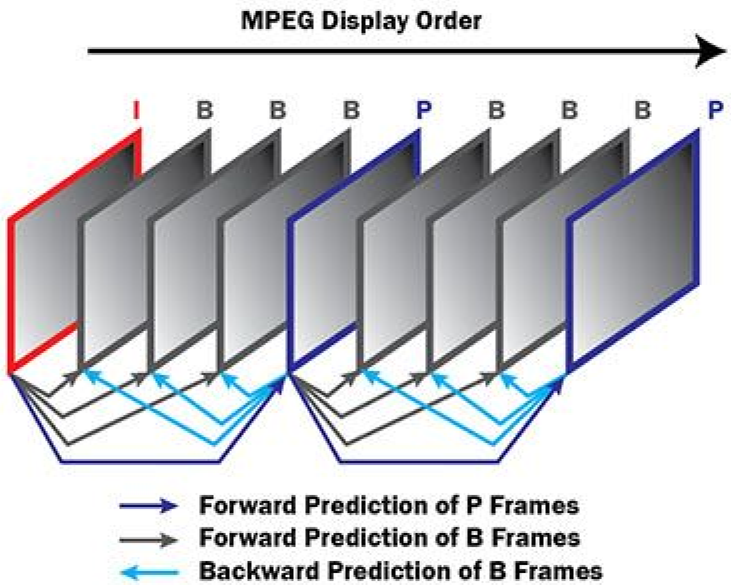

Figure 1: MPEG Frame display order and prediction patterns.[11]

This is because in terms of size, P frames are typically 20 to 70 percent the size of an I frame, where B frames are about 5 to 40 percent the size of an I frame.[13] This size difference is due to how the encoding will ignore much of the frame if it is expected to be the same.

2.1.2 - H.261 / H.263

H.261 is a more recent video compression technology.[12] It also utilizes the idea of intra frames, although somewhat differently than MPEG. It was developed specifically for video conferencing and those applications with small ranges of motion. Many of the minute details of

H.261 are vendor specific, but the general format is the same for all programs. Unlike MPEG technology, H.261 does not encode the differences in motion of the video, but performs calculations determining if motion is detected. Every frame of video is split into groups of blocks, or GOBs. Larger images are split into more GOBs than smaller images. These GOBs are then divided into macroblocks, relating to pixels on a screen. Each of these are split into luminance blocks and spatial color difference blocks. Figures 2a and 2b show the breakdown of these GOBs. Each block (numbered 1 to 33) in Figure 2a is broken up further into the blocks shown in Figure 2b.

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

Figure 2a: Visualization of macroblocks within a GOB for H.261.[12]

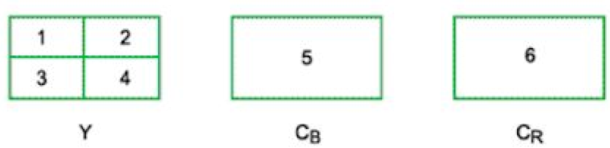

Figure 2b: Visualization of blocks within a macroblock for H.261.[12]

When the decoder looks at the incoming data stream, it compares blocks, and not entire pictures. The blocks are then encoded and decoded, then compared. Comparisons are taken between the two images in a stream. If there is a noticeable difference, the encoder will translate the difference. Because the blocks are broken into separate groups of pixels, the predictive process takes place on this level.

H.263 is very similar to H.261, with some minor adjustments. One is the addition of a PB frame, or two pictures coded as one unit. This frame consists of one predicted picture from the last decoded frame, and a “B frame” (the same concept as in MPEG) predicted from the last predicted image and the image currently being decoded. Figure 3 shows this new type of frame and its “predictive” patterns. This combination allows the frame rate to be increased without increasing bitrate.

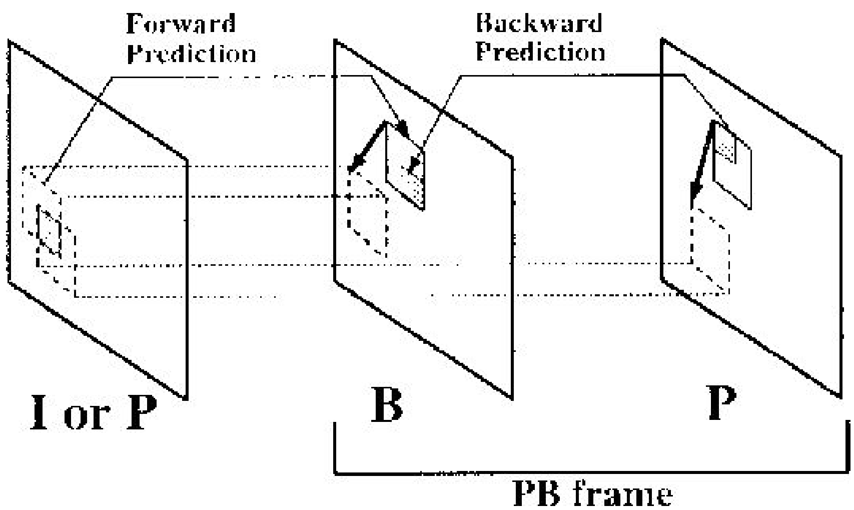

Figure 3: Visual Description of PB frames in the H.263 compression technology.[14]

2.1.3 - Comparison

MPEG can use any image size and has the ability to handle video with a large range of motion.[13] H.261, as it was designed for video conferencing, uses the faster way of encoding to excel at small motion streaming. Additionally, because of the structured blocks, H.261 may have more stability in regards to packet loss vs MPEG. H.261 does not rely on a stream of data and frames but rather comparisons between decoded frames. The loss of an I frame in MPEG could cause a drastic change where in H.263 it can be mitigated. The effect of packet loss in MPEG was shown in the study by Greengrass, Evans, and Begen to cause screen tearing and video artifacts.[15] Figure 4 shows a summary of these two technologies:

Type	MPEG	H.261/H.263
Predictive?	Yes	No
Type of Motion?	Large range of motion	Small Range of motion
Affected by packet loss?	Affects Tearing	Must be investigated

Figure 4: Comparison of MPEG and H.26X series

2.2 - Digital Communication Architectures

Video conferencing and calling technology can use different architectures, which determine how users connect to each other when a call begins. There are 2 main types of video calling architectures: peer to peer connection and client-server connection.

2.2.1 - Peer to Peer Connection

Early services used what is referred to as peer to peer connection. Rather than having each participant connect to a central server, each participant connects directly to their counterpart. There is no central control. All computers act as the client and the “server” at the same time.[16] Figure 5 visualizes this relationship in greater detail. The arrows represent individual data streams.

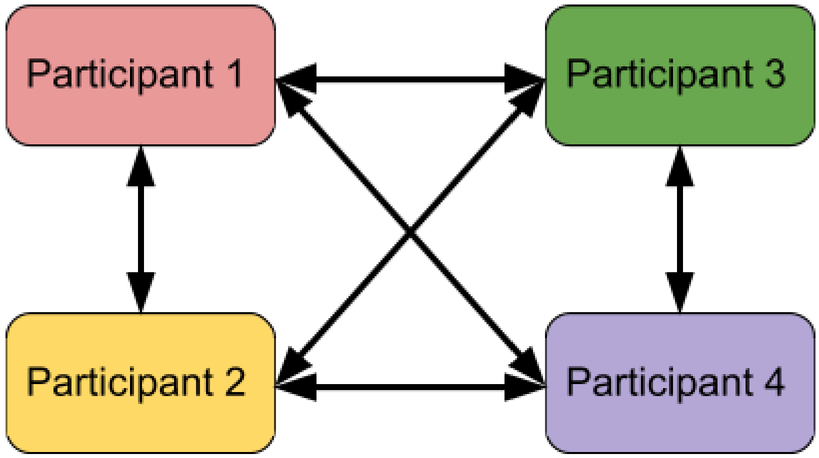

Figure 5: Example of peer to peer based communication systems.

Zoom offers peer to peer connections for their free, lower bitrate calling services.[17] Early video calling services used peer to peer connections due to its cheaper nature. There is no need to pay for server maintenance since the connection is only created when participants start a call. No central computing power is needed, as the users computers do the bulk of the encoding and decoding. However, as seen in Figure 5, this architecture does not scale well past a limited number of users. Additional problems are encountered when dealing with variables such as distance and clients with limited bandwidth.

2.2.1 - Client-Server Connection

Client-Server connections are the most popular today. These are what are used by Zoom for their paid packages.[17] In this architecture, the participants of the call both connect to a central server.[16] The participants do not need to share resources with each other, as the server should have most of the computing power. Each user sends their information to a central connection. This central connection has a single IP, and coordinates the incoming and outgoing information for all the users. In theory, this system is more stable with multiple users. As Figure 6 shows, scalability is not dependent on the individual peers but only on the server.

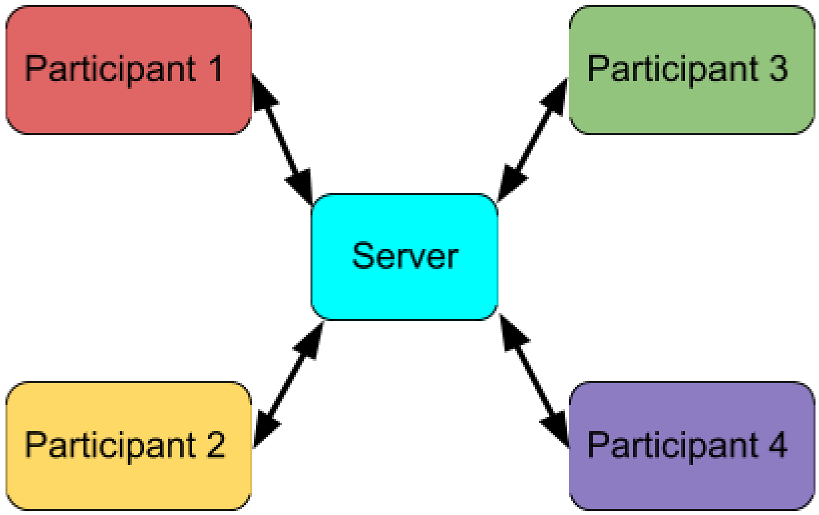

Figure 6: Example of server based communication system.

2.3 - Network Protocols

A protocol, also called an access method, defines how data is exchanged over a network.[18] Every protocol has its own method for formatting, receiving, compressing, sending, and error-checking data. Three protocols are referenced in this paper: TCP, UDP, and MQTT.

2.3.1 - TCP

Transmission Control Protocol (TCP) is one of the most widely used protocols. TCP is used for most methods of communication and text display, including email and many websites.[19] The Internet uses a combination of TCP and IP (Internet Protocol). TCP falls under one of the four layers of the protocol stack, which a computer needs to communicate on the Internet.[20] TCP is the layer responsible for transmitting data between communicating devices.

TCP is connection-oriented, which means it must establish a connection with another device before it starts sending data. It is built with error-checking, meaning TCP will resend a dropped packet should that packet not make it to its intended destination, as shown in Figure 7. This can lead to large delays, since TCP will not send another packet until the dropped packet is received.

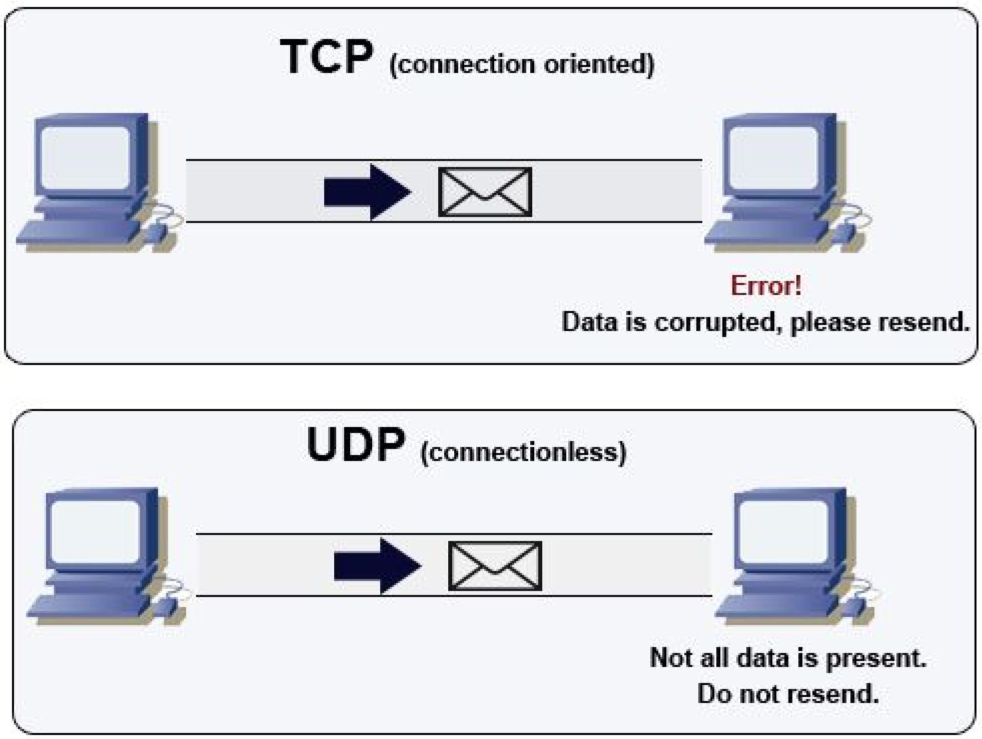

Figure 7: Shows the difference between how TCP and UDP sends packets and responds to packet loss.[21]

2.3.2 - UDP

User Datagram Protocol (UDP) is another protocol often used by Internet applications. UDP is similar to TCP when running over IP, except it does ensure every packet is delivered.[22] If a packet is dropped in a UDP connection, UDP will not try and resend the packet: it will just continue to send packets as if nothing happened, as illustrated in Figure 7. Dropped packets cause a drop in quality, since not all of the data is received, but there is no delay since the stream of packets is not held up waiting for retransmission. UDP sees a lot of use in video communications and online gaming because dropping packets results in a better interactive experience than the extra delay of sending packets again.

2.3.3 - MQTT

Message Queuing Telemetry Transport (MQTT) is a text-exclusive protocol.[23] MQTT runs over other bi-directional protocols, such as TCP. It utilizes a publish/subscribe message pattern that allows for one user to send a message to many recipients. Each message can use one of three qualities of delivery. “At most once” messages are sent in the lowest quality. Messages can be dropped. “At least once” messages are sent until the device sending the message is sure that the recipient has gotten that message, but the recipient could receive duplicates. “Exactly once” messages are made sure that the message is received exactly once. This is important in billing systems.

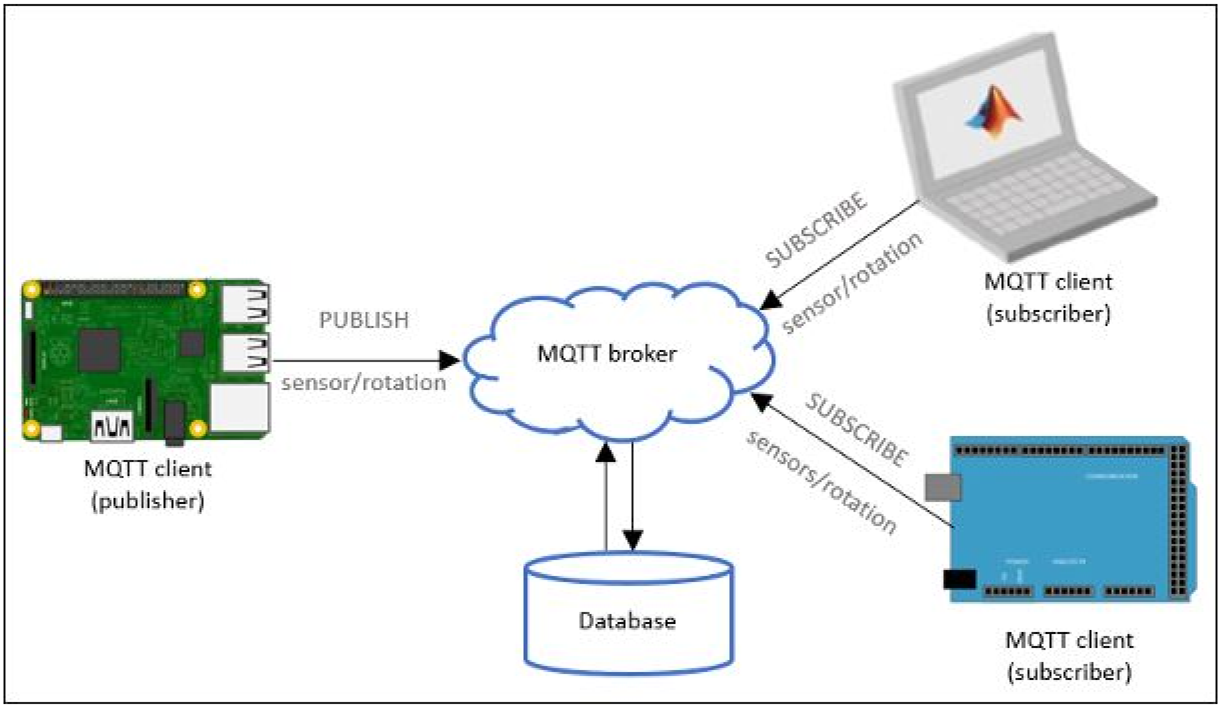

Figure 8: Shows how messages are sent and received using MQTT.[24]

2.4 - Tools

Various types of tools are used in our experiments with video conferencing. We need to manipulate the network conditions and measure the effects on packets and bitrate.

2.4.1 - Wireshark

Wireshark is an open source packet analyzer (often called a packet sniffer) used for network analysis and troubleshooting.[25] Wireshark captures packets by putting the users network card into promiscuous mode which makes the computer accept all packets. Wireshark can read network traffic from the Ethernet, WiFi, PPP/HDLC, ATM, Bluetooth, USB, Token Ring, Frame Relay, and FDDI.[25]

Wireshark has two default filters with their own syntax.[26] The capture filter is set before the packet capturing process begins, choosing which packets to capture. The display filter is used to hide irrelevant packets, and after capture allows the user to focus on specific packets. Both of these tools can filter by ports, protocols, addresses and other network properties.

Wireshark has many post capture analysis tools. Wireshark shows the content of the packet as well as packet length, source port, and destination port. Wireshark also has many graphical tools, the most important for our experiment being the I/O graph. The I/O graph can be configured to show different packet properties over time.

2.4.2 - Clumsy

Clumsy is a program that can simulate problems that can occur in network connections by capturing, distorting then releasing packets.[27] Clumsy can drop random packets, throttle traffic, duplicate/clone packets, randomly rearranging the order of packets, and tamper with the packet itself. We use Clumsy to simulate latency and to drop packets on a percentage basis.

3 - Methodology

Understanding the effects of bandwidth and packet loss on recent video conferencing technologies is key to improving the user experience. To gather data on these effects, we have developed a methodology for both quantitative and qualitative analysis of two video conferencing services. This chapter describes why these services have been chosen, as well as the tools and solutions used to conduct this experiment. This chapter also explains the process behind the collection of data, both from the hardware/software itself as well as collecting users opinions of the calling experiences.

3.1 - Video Conferencing Services

There are numerous different video conferencing tools available today. Some are used in trusted businesses, while others are targeted towards the average consumer. We will focus our study towards businesses.

3.1.1 - Zoom

Zoom is a video conferencing company founded in 2011 and has grown into a multi-billion dollar company since.[28] Zoom is a service designed more for businesses than average home consumers, with the ability to have large amounts of callers and the creation of "meetings."[3] Video calls in Zoom can have up to 1000 people, or participants, and up to 49 can share their screens. Zoom markets its tools to facilitate engagement between both customers and employees, as well as provide a platform for meetings and conferences.[4]

3.1.2 - Microsoft Teams

Microsoft Teams is the follow-up to Microsoft's" Skype For Business". It is available in Windows, MacOS, Linux, IOS, and Android operating systems. Microsoft Teams has a video call feature that can include up to 80 participants. Teams can do file storage as well as file collaboration, it also has a calendar application with reminders.

3.2 - Digital Tools

Our experiment focuses on user response to video conferencing quality under different network conditions. To be able to conduct this experiment we required tools that measure network statistics and to manipulate network conditions. We use the program Clumsy(V2) to alter these properties.[27] Clumsy allows us to add network latency, packet drop chance, and throttle network data capacity. To measure packet loss and factors contributing to video and audio quality we used Wireshark and Fraps. Wireshark is a network sniffer that can capture and measure network data[25]. It has various tools including packet analysis over time and TCP errors over time. Fraps can be used to measure the frames per second of the videos.

3.3 - Parameters

We use latency, bandwidth, and packet loss as parameters in our study. Effects on quality of experience (QoE), frames per second (fps), and packet data are collected.

3.3.1 - Latency

Latency is the total amount of time taken for data to be sent from one machine to another, received, and then sent back again.[29] Network devices will often send a small amount of data and wait for an acknowledgment that the data is received, and then send again. A higher latency will result in a longer delay for information being sent and received. Latency is typically measured in milliseconds (ms).

3.3.2 - Packets

Packets are small pieces of data sent to another destination over a network.[30] Network devices will break up large data in smaller packets to send the data to another device,sometimes by multiple routes. The device receiving the packets reassembles the packets back into the original data. If some packets are lost, devices will still reassemble packets together, but the data received will be incomplete. Figure 9 shows how packets are disassembled and reassembled.[31]

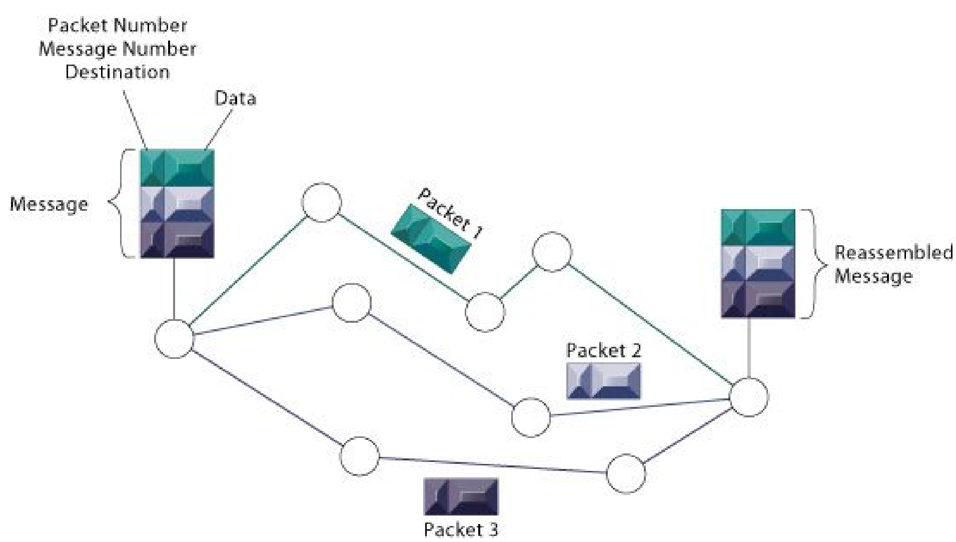

Figure 9: Packet disassembling and reassembling.[31]

3.4 - Laptop

Laptops used were Dell Latitude E5450 running Microsoft Windows 10. Figure 10 shows the specific specifications for this model of laptop. In order to avoid any performance drops, the laptops were plugged in during all tests.

Processor	Graphics Card	Resolution	Physical Memory
Intel Core i7-5600U CPU @ 2.60GHz	Intel HD Graphics 5500 1GB RAM	1920x1080x60hertz	8GB

Figure 10: Laptop Specifications.

3.5 - Experimental Design

We divide our experiments into two sections. The first is a user study, where subjects use the software and rate their experience under different network conditions. The second is benchmarking and collecting network traffic data. Both sections compare the same network condition,which allows comparison between technical data and user experience.

3.5.1 - User Study

The user study consisted of a video conference session with three participants in each. These participants were WPI students. The participants were told to go to a preselected classroom, and filled out informed consent papers. The students were then directed to one of three laptops. All three laptops were the same model, and were equipped with identical noise reduction earbuds. All three laptops were connected to a wired ethernet connection, by means of a network splitter and identical ethernet cords. Participants were spread out as much as possible, to avoid hearing each other in any way except through the earbuds.

On each of the laptops were loaded the Zoom application, the Teams application, and the program Clumsy. The investigating team had before hand written a script, which controlled Clumsy. This was also loaded on each individual computer. The team had made sure that aside from the video conferencing software being tested, all other applications (aside from Clumsy) were not running and were shut down. Before the call began, users were asked to fill out a short demographics sheet. This included age, major, gender, and their experience with various types of wireless communication. These include experience with video streaming, voice over IP services, and video calling services.

Each call began with a short, semi scripted conversation where the users introduced themselves. This included their name, major, year of graduation, and a fun fact if they desired. The subjects were then told to rate the interaction based on previous experience with video calling. The users then participated in small joke “rounds”. Each person would say a short scripted joke, and after each participant had told their respective joke a short discussion would follow. After each activity, users were asked questions on how the video and audio quality was affected. On a scale of one to five, where 5 is the better experience, they were asked to rate the questions listed in Figure 11.

The presence of audio glitches
The presence of video glitches
The video experience
The audio experience
The overall experience of the call

Figure 11: The questions participants filled out after each "round."

Each of these "rounds" corresponded to different network conditions, which were changed using the .bat script from before. The levels tested are shown in Figure 12. This order was randomly generated to avoid the participants from noticing a pattern.

Rounds	1	2	3	4	5	6	7	8	9	10
Packet Loss(%)	0	2	20	0	20	0	2	0	2	2
Latency(ms)	100	100	0	0	200	200	0	50	200	50

Figure 12: List of latency and packet drop levels.

3.5.2 - Benchmarking

For benchmarking data was collected on specific network attributes. All three laptops were the same model, and were equipped with identical noise reduction earbuds. All three laptops were connected to a wired ethernet connection, by means of a network splitter and identical ethernet cords. On each of the laptops were loaded the Zoom application, the Team's application, and the program Clumsy. The same .bat script was used to control Clumsy. This was also loaded on each individual computer. The team had made sure that aside from the video conferencing software being tested, all other applications (aside from Clumsy) were not running and were shut down. Going through the same levels of packet drop and latency, Wireshark was used to record packet and byte data for both video conferencing software. The team gathered tables of data on bytes per second, packets per second, packet length, and packet size. This was done with the same network setup as with the user study. Each value of network manipulation was measured for about 30 to 45 seconds and saved for later analysis.

4 - Analysis

This chapter reports and analyzes the data collected during our benchmarking and our user study. Section 4.1 is the analysis of Zoom and Microsoft Teams benchmarking data on packets in different network conditions. Section 4.2 is the analysis of our user study. Section 4.3 is a summary analysis and how the benchmarking data relates to the user data.

4.1 - Benchmark Results

We repeated the user study ourselves with wireshark capturing all network traffic. We used the same network conditions as the user study and logged all packet traffic for 30-60 seconds per round. There were a total of ten rounds with different parameters as shown in Figure 12 on chapter 3 (Methodology). All three laptops measured packet traffic. Packet traffic measured was packets per second, bytes per second, and packet length. Bytes per second was measured during benchmarking as it more accurately represents data usage compared to packets per second.

The graphs were generated with Microsoft Excel. Every combination of packet drop and latency had more than 30 observations, so we used a normal distribution to generate 95% confidence intervals for each mean. In most cases, the confidence intervals were incredibly small, sometimes smaller than 5,000 bytes.

4.1.1 - Zoom

Figure 13 shows the kilobytes per second versus the packet drop chance. The y-axis is kilobytes per second and the x-axis is the packet drop chance. The blue line represents 0 milliseconds of latency while the orange line represents 200 milliseconds of latency. Packet drop does have a notable change in bytes per second as seen in Figure 13. The rate of bytes per second increases from about 328,000 at no packet drop chance to about 336,000 at 2% packet drop chance and about 407,000 at 20% packet drop chance. This equals an increase of 2.36% and 24.0% respectively. We can observe that for Zoom packet drop percentage causes a linear, one-to-one increase in bytes per second.

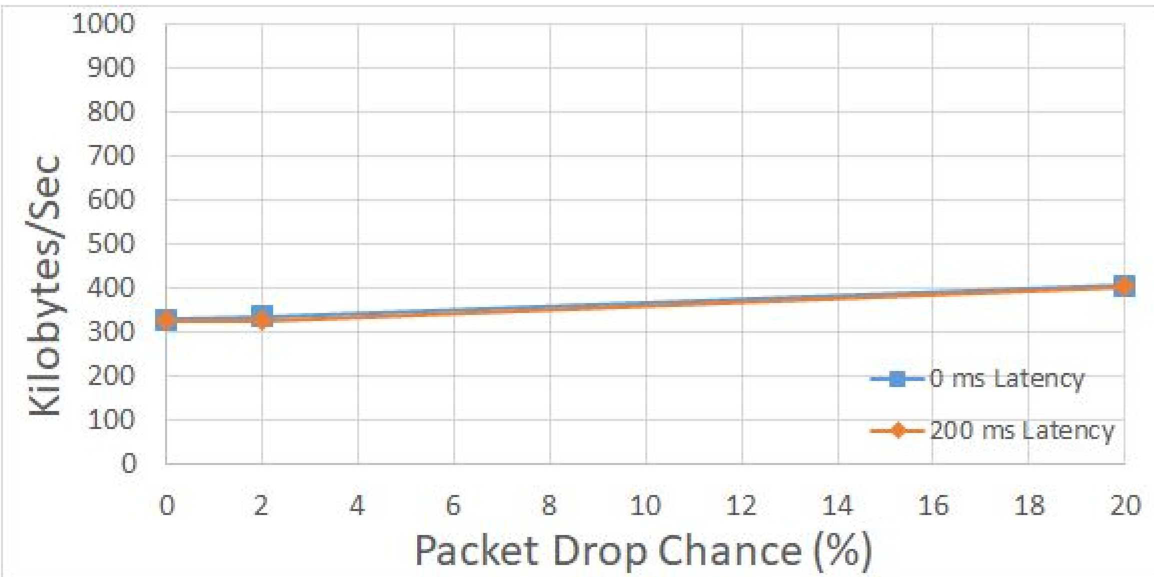

Figure 13: Graph of bytes per second based on packet drop chance for Zoom.

Figure 14 shows the kilobytes per second versus the packet drop chance. The x-axis is the latency in milliseconds and the y-axis is kilobytes per second. The blue line represents 0% packet drop while the orange line represents 2% packet drop. With Zoom, changes in latency have no significant statistical effect on bytes per second as can be seen in Figure 14.

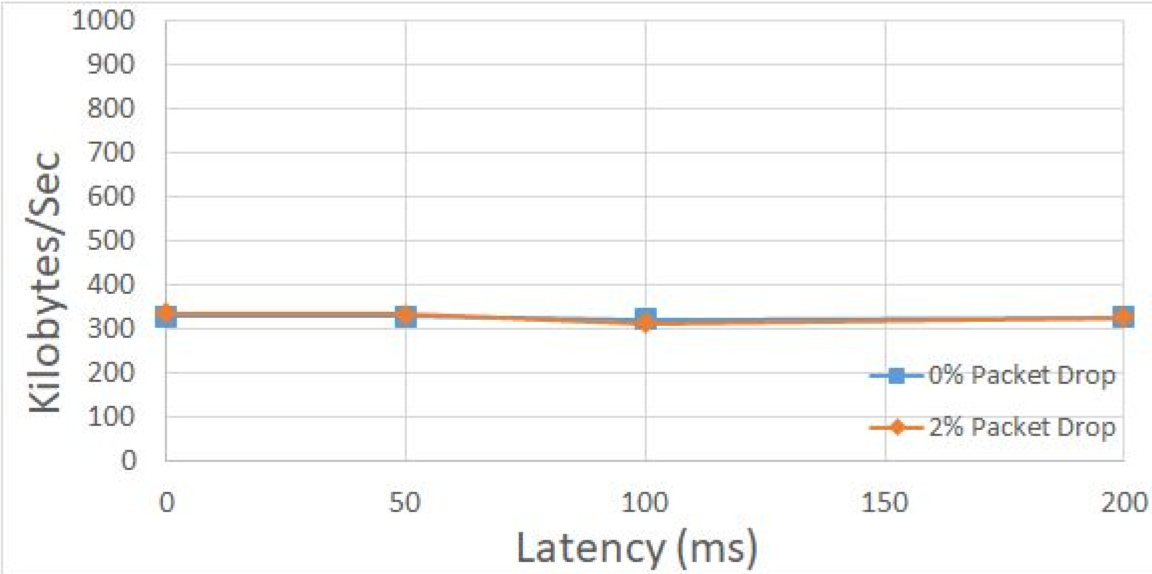

Figure 14: Graph of bytes per second based on latency levels for Zoom.

4.1.2 - Microsoft Teams

Figure 15 shows the kilobytes per second versus the packet drop chance. The x-axis is the packet drop chance and the y-axis is kilobytes per second. The blue line represents 0 milliseconds of latency while the orange line represents 200 milliseconds of latency. In Figure 15 the rate of bytes per second dropped by approximately 35% at all levels of packet drop chance measured from no latency to 200 milliseconds (ms) of latency. The total amount of bytes per second also dropped by approximately 35% between 0% packet drop chance and 20% packet drop chance at both levels of latency.

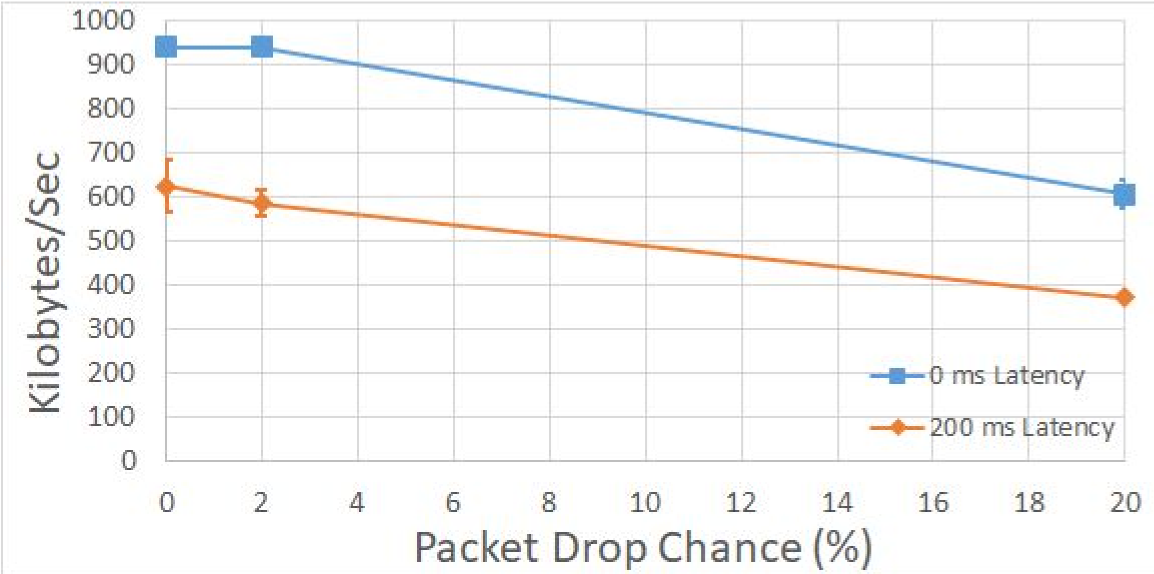

Figure 15: Graph of bytes per second based on packet drop chance for Microsoft Teams.

Figure 14 shows the kilobytes per second versus the packet drop chance. The x-axis is the latency in milliseconds and the y-axis is kilobytes per second. The blue line represents 0% packet drop while the orange line represents 2% packet drop. Figure 16 shows that the bytes per second rate experienced a similar decrease as latency increased, but 50 ms of latency with 2% packet drop chance drops lower than 200 ms of latency. Without this point, there appears to be a linear decrease in bytes per second.

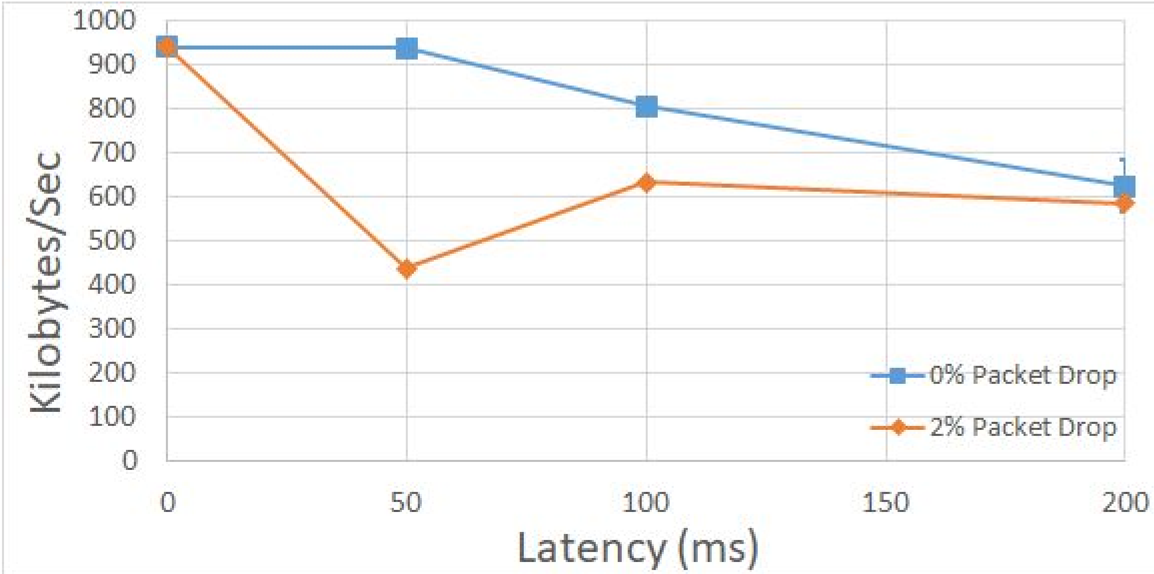

Figure 16: Graph of bytes per second based on latency levels for Microsoft Teams.

4.1.3 - Zoom/Teams Comparison

While Zoom showed a small increase in total bytes per second when increasing packet drop chance, Teams showed a large decrease in total bytes per second. However, Teams still sent substantially more bytes per second than Zoom.

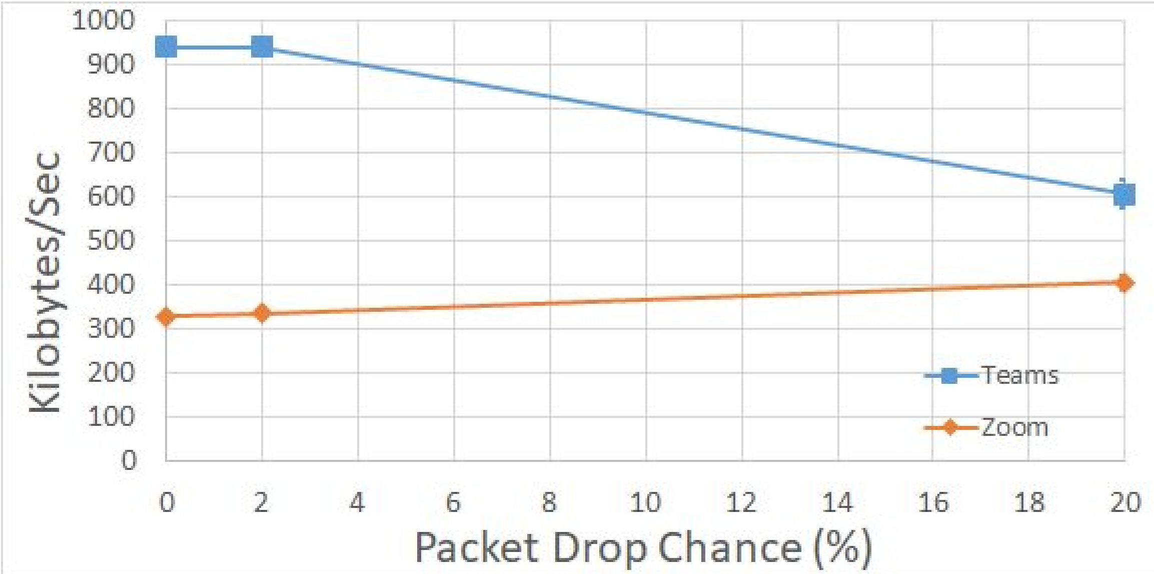

Figure 17: Comparison between Zoom and Teams bytes per second with no latency.

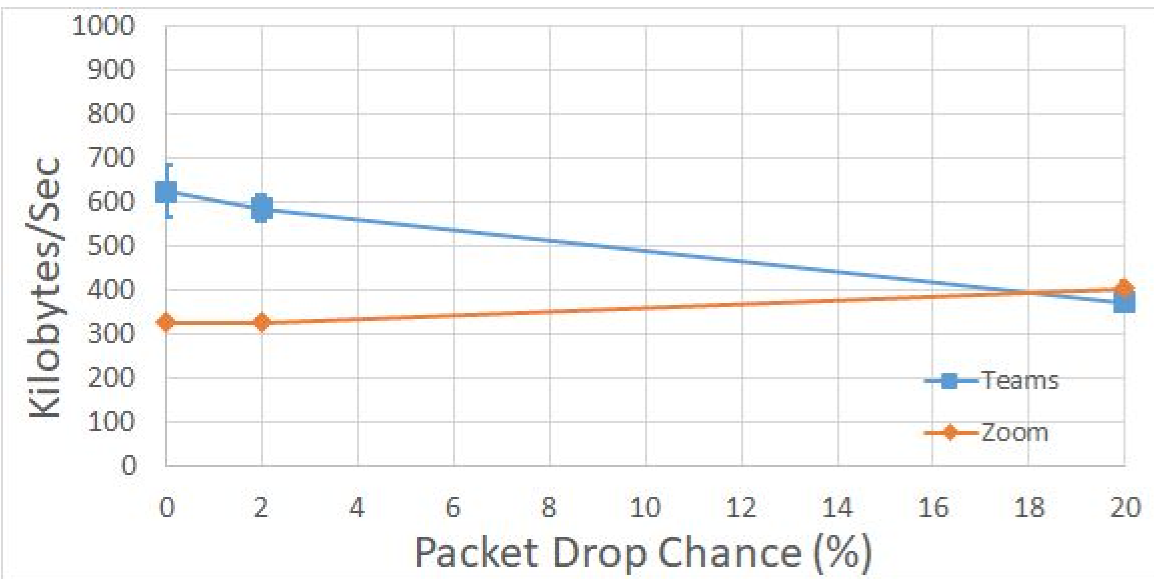

Figure 18: Comparison between Zoom and Teams bytes per second with 200 ms latency.

In Figure 18, Zoom still experienced the same increase as in Figure 17, and Teams also experienced the same decrease as Figure 17. However, Teams' bytes per second rate is affected by latency. Teams appears to have the same rate of bytes per second as Zoom when the network has 200 ms of latency and 20% chance of packet drop.

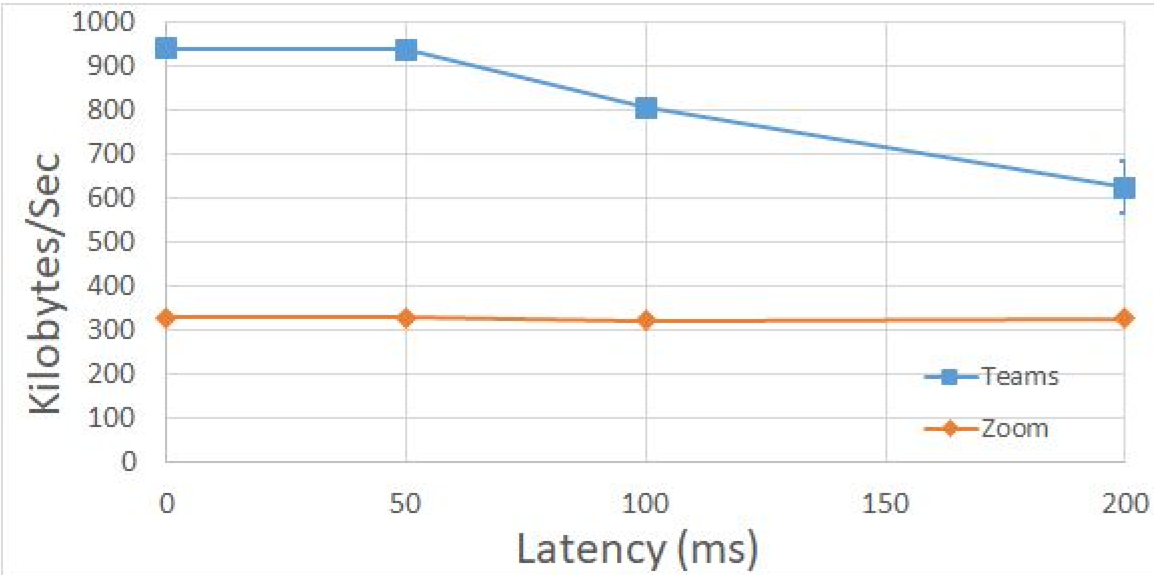

Figure 19: Comparison between Zoom and Teams bytes per second with 0% packet drop chance.

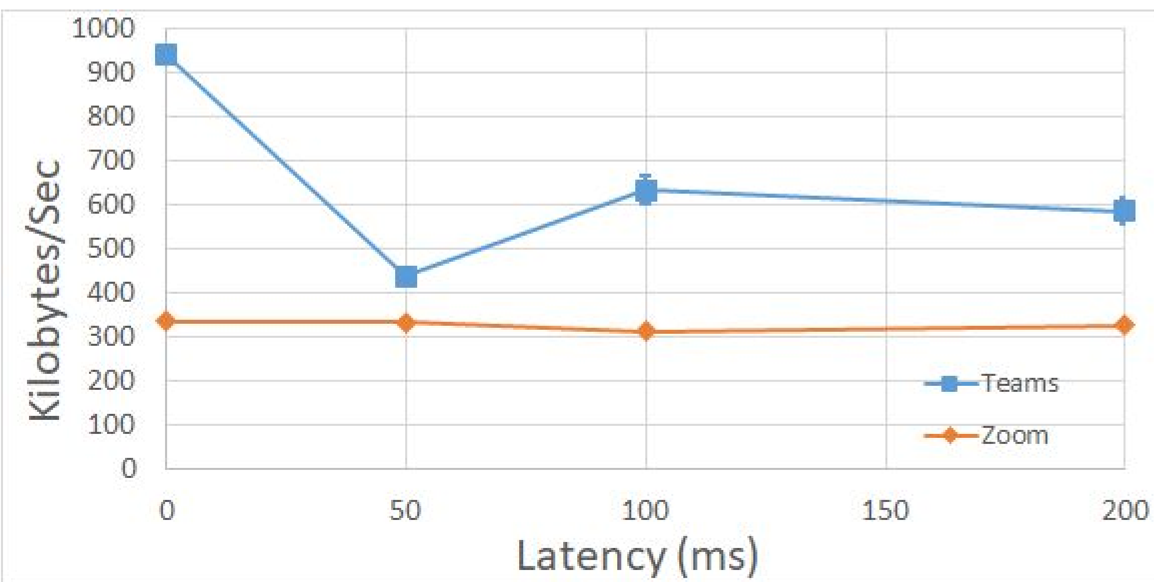

Figure 20: Comparison between Zoom and Teams bytes per second with 2% packet drop chance.

4.2 - User Study Results

This section analyzes the user data. Both Zoom and Teams are analyzed individually and compared, along with an examination of the demographics. We had a total of 22 participants for our study. There were a total of ten rounds with different parameters as shown in Figure 12.

Since we only had 22 participants in our study a t-distribution with 21 degrees of freedom was used to determine the 95% confidence intervals for each mean.

4.2.1 - Demographics

Figures 21, 22, and 23 gives an overview of the demographics in our study. Our participant pool is almost 50/50 with male and female participants. In general, the participants are young, college aged, and generally experienced with video/audio communication software. The participants are mostly engineers and computer scientists.

Reported Age and Gender	**18**	**19**	**20**	**21**	**Male**	**Female**
Number of Participants	3	9	6	3	11	9

Figure 21: Table of reported demographic information for participants.

Rating	**Video Streaming Experience**	**Voice over IP Experience**	**Video Calling Experience**	**Zoom Rating**	**Teams Rating**
1	2	2	0	0	1
2	1	0	3	2	5
3	6	5	2	12	12
4	6	5	6	6	4
5	7	10	11	2	0

Figure 22: Table of participant responses to background questionnaire. 5 represents more experience/better experience, where 1 represents less experience/worse experience.

The participants initially rated the experience of Teams and Zoom during an introduction phase. There was no network manipulation during this phase.

Reported Major	Number of Participants
CS & Interactive Media and Game Development (IMGD)	6
Mechanical Engineering	3
Chemical Engineering	2
Computer Science (CS)	1
Electrical and Computer Engineering (ECE)	1
Biotech & Biochem	1
CS & Psychology	1
Aerospace Engineering	1
CS & ECE	1

Figure 23: Table of majors for the participants.

4.2.2 - Zoom

Figure 24 shows the average rating versus the packet drop chance. The y-axis is average ratings and the x-axis is the packet drop chance. The blue line represents 0 milliseconds of latency while the orange line represents 200 milliseconds of latency. Zoom does not seem to have significant drop in ratings as packet drop chance increases. The drops in Figure 24 are within the confidence intervals.

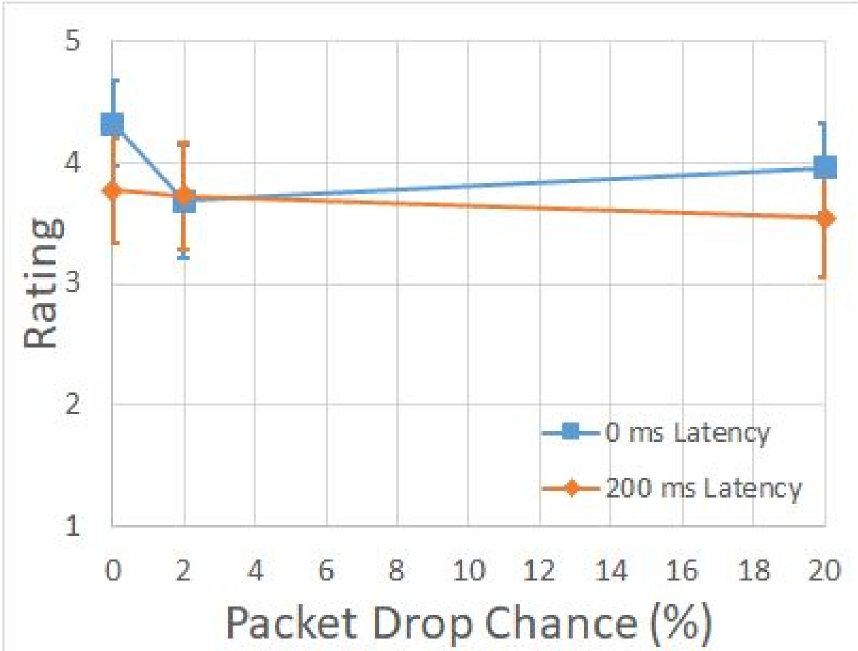

Figure 24: Graph of average user ratings based on latency for Zoom.

Figure 25 shows the average rating versus the latency in ms. The y-axis is average ratings and the x-axis is the latency. The blue line represents 0% packet drop while the orange line represents 0% packet drop. In Figure 25 all the confidence intervals overlap so it is difficult to see a trend other than latency has no effect on rating. The only potential noteworthy change is between 0% drop/0 ms delay and 20%/200 ms where the confidence intervals barely overlap and the average rating drops by almost 1 point.

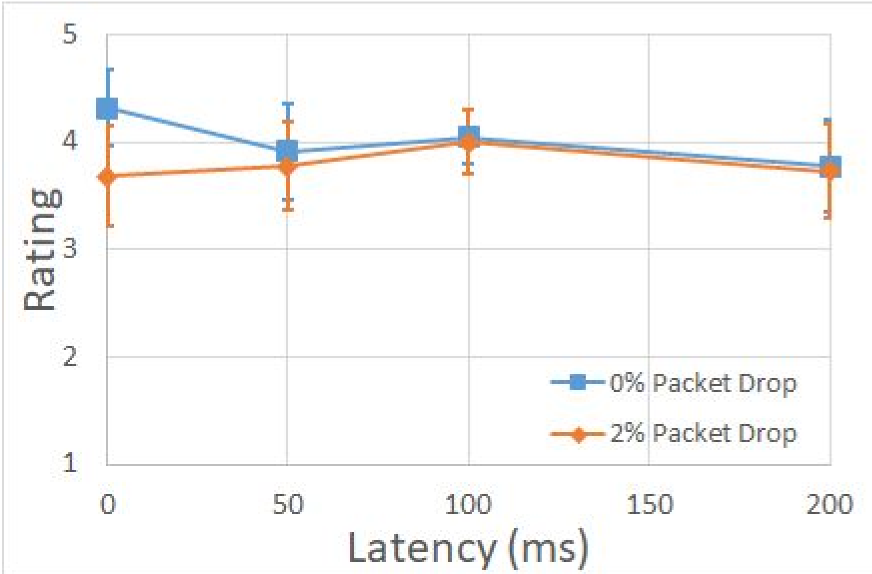

Figure 25: Graph of average user rating based on packet drop chance for Zoom.

4.2.3 - Microsoft Teams

Figure 26 shows the average rating versus the packet drop chance. The y-axis is average ratings and the x-axis is the packet drop chance. The blue line represents 0 milliseconds (ms) of latency while the orange line represents 200 milliseconds of latency. While the confidence intervals barely overlap, we can observe an almost 1 point difference between 0% packet drop chance and 20% packet drop chance for 0 ms of latency in Figure 26. The average ratings of 0% packet drop chance and 0 ms latency, and 20% packet drop chance and 200 ms latency differ by more than 1 point. The average ratings of 0 ms of latency and 200 ms of latency at 0% packet drop chance also differ by 1 point. The confidence intervals barely overlap.

Figure 27 shows the average rating versus the latency in ms. The y-axis is average ratings and the x-axis is the latency. The blue line represents 0% packet drop while the orange line represents 0% packet drop. Figure 27 suggests that the user experiences does not seem to be affected by changes in latency if one considers the small overlap in confidence intervals at 0ms and 200ms for 0% packet drop. Ignoring the small overlap there is around an 0.8 rating drop from 0ms to 200ms. The only potential outlier that might be considered is the difference between 0 ms and 200 ms latency at 0% packet drop chance mentioned above where there only appears to be a slight overlap in between the confidence intervals.

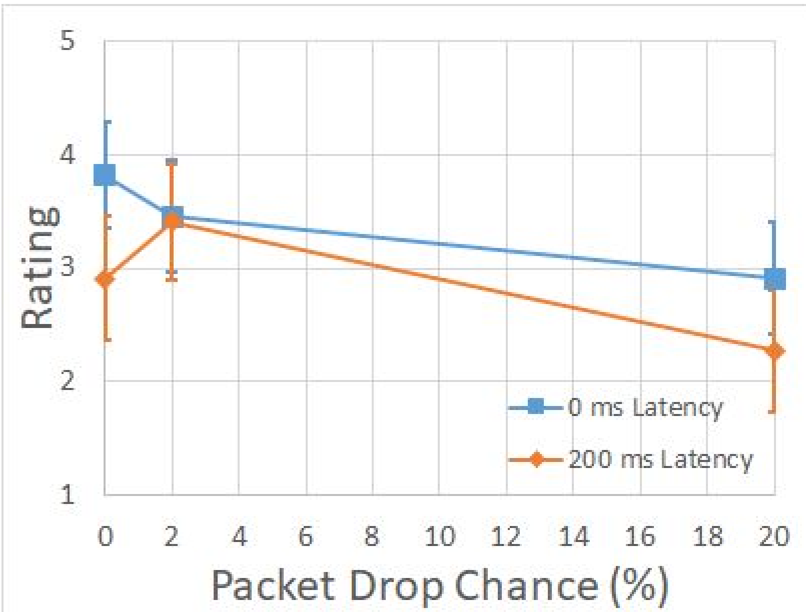

Figure 26: Graph of average user rating based on latency for Microsoft Teams.

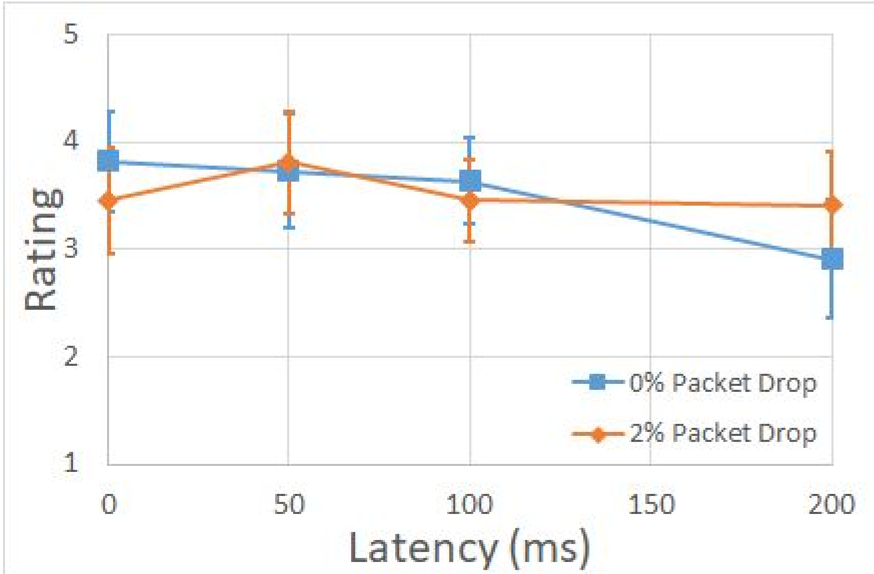

Figure 27: Graph of average user rating based on packet drop chance for Microsoft Teams.

4.2.4 - Zoom Teams Comparison

Without latency, Teams and Zoom ratings appear to decrease between 0% and 20% packet drop chance, as shown in Figure 28. Teams ratings decreased between 2% and 20% packet drop chance, but Zoom saw an increase in average rating. Figure 29 shows that with latency, the average Teams rating for 2% packet drop chance was higher than for 0%, but Zoom's average rating stayed roughly the same. These drops are very slight, and the confidence intervals overlap, so there is no significant change in the data.

Zoom's average rating saw a small decrease as latency increased at 0% packet drop chance, but Teams saw a much larger decrease in ratings between 100 ms and 200 ms of latency (Figure 30). Figure 31 shows at 2% packet drop chance, the ratings decrease from 0% packet drop chance, but then behave differently at 20% packet drop chance.

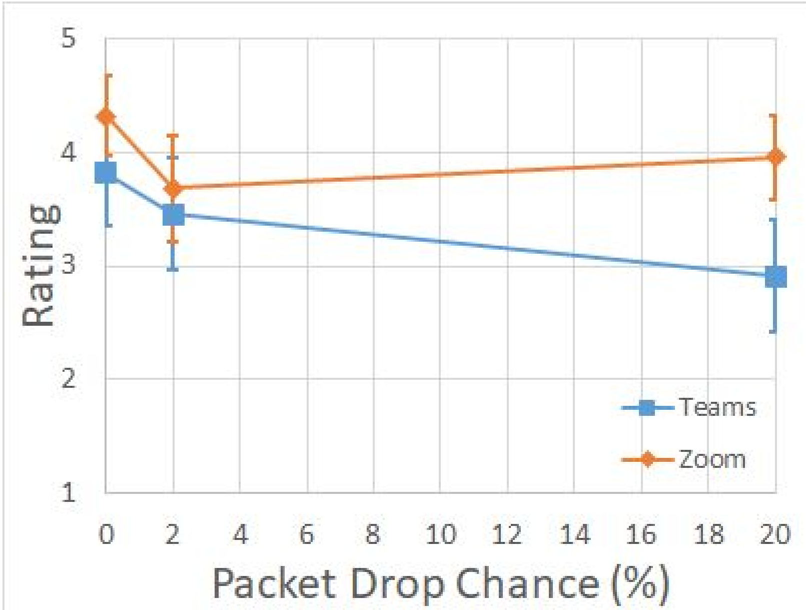

Figure 28: Graph of comparison between Zoom and Teams user ratings with no latency.

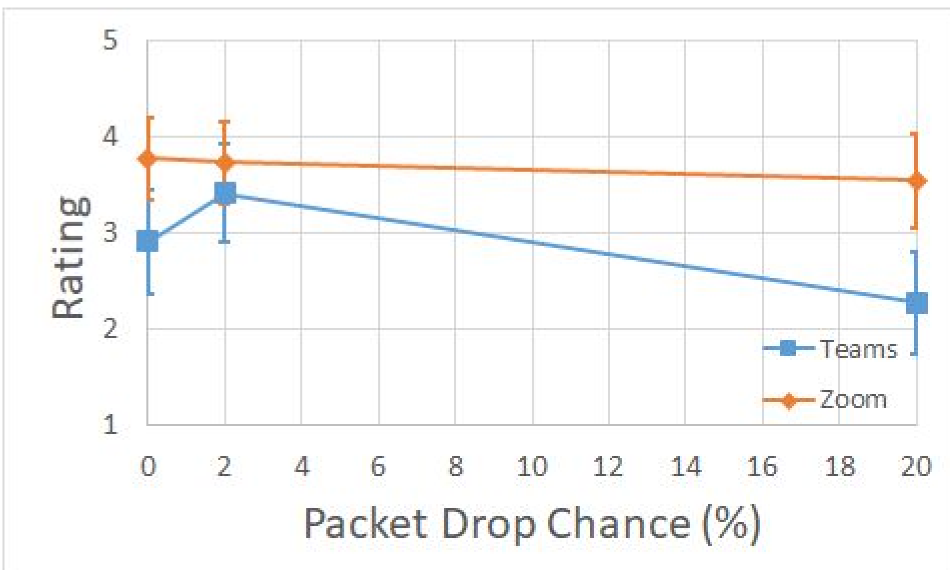

Figure 29: Graph of comparison between Zoom and Teams user ratings with 200 ms latency.

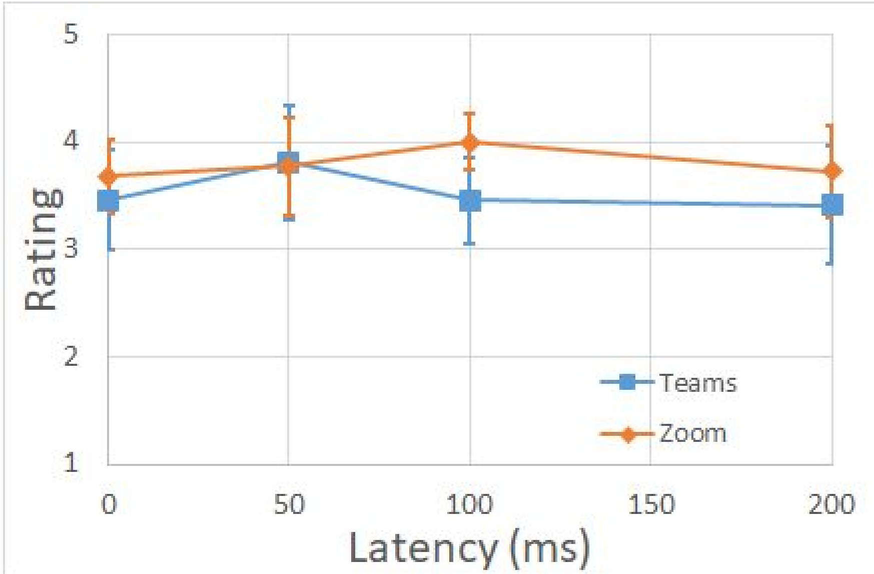

Figure 30: Graph of comparison between Zoom and Teams user ratings with 0% packet drop chance.

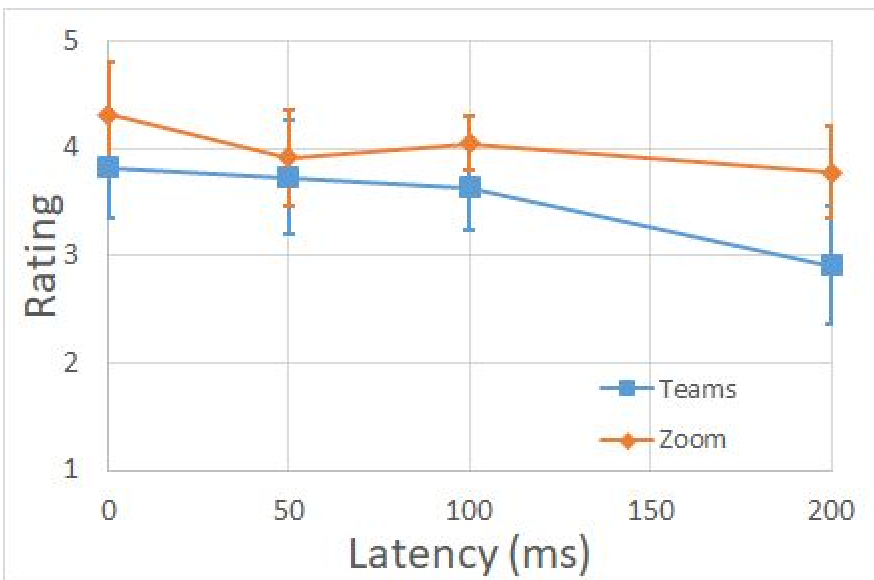

Figure 31: Graph of comparison between Zoom and Teams user ratings with 2% packet drop chance.

4.3 - Overall Results

Based on the margins of error, neither Teams nor Zoom have a statistically significant difference in average rating due to latency or packet drop chance. The confidence intervals had overlaps between different levels of packet drop chance and latency. Despite the lack of significant difference in rating, there were practical differences in Teams due to packet drop chance. At both levels of latency, the maximum and minimum average ratings differed by about 1 point. Since our rating scale was from 1 to 5, 1 point is a significant difference. For different levels of latency maximum and minimum average ratings at 0% packet drop chance differed by about 1 point, but only differed by about 0.4 points at 2% packet drop chance. This is less than 10% difference between maximum and minimum average ratings.

There was a statistically significant difference in average rating between Teams and Zoom. At 20% packet drop chance, Zoom's average rating was about 1.4 points higher than Teams' average rating for both levels of latency. The confidence intervals do not overlap. At 0% packet drop chance, Zoom's average rating was about 0.5 points higher than Teams' with no latency and about 0.9 points higher than Teams' with 200 ms of latency. There is a very slight overlap between confidence intervals, but since the amount of overlap is so small, a significant difference can be observed.

5 - Conclusion

As companies globalize, the demand for quality video conferencing increases. Despite this, there are very few studies done on newer video conferencing systems. We focused our study on Microsoft Teams and Zoom, fairly recent video conferencing systems aimed at the business sector.

Our evaluations were composed of two parts: a technical benchmark and subjective user study. Both of these consisted of 10 rounds per video conferencing system. Each round was a combination of packet drops and/or as latency on three identical computers for a minute per round. A tool called Clumsy was used to change network conditions. The benchmark used Wireshark to log packets per second, bytes per second, and packet lengths. The user study consisted of three participants telling brief jokes every round and then rating their subjective experience from 1 to 5.

Upon analysing the byterate from Wireshark data, it was observed that Zoom stayed mostly constant. Zoom's byterate increased roughly linearly with packet drop chance increase. Latency had no effect on Zoom's byterate. Teams had large changes in byterate due to packet drop chance and latency, byterate significantly decreased when latency increased and when packet drop chance increased. The highest levels of packet drop chance and latency resulted in 35% drops in byterate. Under normal network conditions, Teams' rate was nearly triple Zoom's, but at the worst conditions, Teams' rate was slightly lower than Zoom's.

From our user study, we saw that ratings stayed fairly consistent for both services separately. However, Teams had a larger decrease in ratings than Zoom did when subjected to large levels of packet drop. Increasing the latency resulted in similar decreases in average rating for Zoom and Teams. These changes were not statistically significant.

We found that Zoom had statistically higher average ratings than Teams. At high levels of packet drop chance, Zoom's average rating was at least 1 point higher than Teams' average rating. Our scale was out of 5 points, so 1 point is a significant difference. At lower levels of packet drop chance, Zoom's average rating was about half a point higher than Teams'. Zoom's and Teams' average ratings were affected by latency similarly.

6 - Future Work

This study looked only at latency and packet drop. In future work, more network variables could be experimented with in addition to latency and packet drop chance. Bandwidth in particular may yield interesting results. This would involve using a tool to restrict the bandwidth allowed and then gather bitrate data using Wireshark.

More levels of packet drop chance and latency could be looked at. Our study only had three levels of packet drop chance and four levels of latency (see Figure 12 for the complete list of values). Getting more levels of packet drop, in between the 2% and 20% tested, may yield data to draw more concrete conclusions.

We were also unable to capture frames per second (FPS) data. We can extrapolate from our user study and from our benchmarks the relationship between the changes in the network and user experience. Having FPS data would show a technical change in how video is viewed to compare with as well.

Another improvement that could be made would involve an expanded user study. We designed our user study to simulate natural, short conversations, while giving the participants time to observe changes. A user study that would be designed over a longer period of time could simulate other conversation types and allow subjects to better report how they experienced the call. This could also be improved by running the study longer and increasing the sample size obtained.

References

1. Business Matters. "The History of Video Conferencing." *Business Matters*, 8 Jan. 2015, https://www.bmmagazine.co.uk/tech/history-video-conferencing/.
2. ReportBuyer. "Global Video Conferencing Market Analysis, Forecast to 2023." PR *Newswire: Press Release Distribution, Targeting, Monitoring and Marketing*, 11 Feb. 2019, https://www.prnewswire.com/news-releases/global-video-conferencing-market-analysis-forecast-to-2023-300793076.html.
3. "Video Conferencing, Web Conferencing, Webinars, Screen Sharing." *Zoom Video*, https://zoom.us/ent?zcid=3172.
4. "Compare Portal Smart Video Calling Devices." *Portal from Facebook*, https://portal.facebook.com/products/.
5. "Video Conferencing, Web Conferencing, Webinars, Screen Sharing." *Zoom Video*, https://zoom.us/government.
6. Novet, Jordan. "Zoom Has Added More Videoconferencing Users This Year than in All of 2019 Thanks to Coronavirus, Bernstein Says." *CNBC*, CNBC, 26 Feb. 2020, www.cnbc.com/2020/02/26/zoom-has-added-more-users-so-far-this-year-than-in-2019-bernstein.html.
7. Anderson, Kareem. "Microsoft Teams for Education Adds Assignments and Grading Features." *OnMSFT.com*, 11 May 2018, www.onmsft.com/news/microsoft-teams-for-education-adds-assignments-and-grading-features.

8. Chan, Rosalie. “Microsoft Teams Just Hit 20 Million Daily Active Users, Beating Its Rival Slack Once Again.” *Business Insider*, Business Insider, 19 Nov. 2019, www.businessinsider.com/microsoft-teams-20-million-daily-active-users-slack-2019-11.
9. Zhang, Xinggong, et al. “Profiling Skype Video Calls: Rate Control and Video Quality.” *2012 Proceedings IEEE INFOCOM*, 2012, doi:10.1109/infcom.2012.6195805.
10. Sat, Batu, and Benjamin Wah. “Analysis and Evaluation of the Skype and Google-Talk Voip Systems.” *2006 IEEE International Conference on Multimedia and Expo*, 2006, doi:10.1109/icme.2006.262681.
11. “Streaming Video and Compression.” *RGB Spectrum*, https://www.rgb.com/streaming-video-and-compression.
12. Davis, Andrew. “An Overview of Video Compression Algorithms.” *EETimes*, EE Times, 3 Feb. 1998, https://www.eetimes.com/document.asp?doc_id=1275884#.
13. Greengrass, Jason, et al. “Not All Packets Are Equal, Part I: Streaming Video Coding and SLA Requirements.” *IEEE Internet Computing*, vol. 13, no. 1, 2009, pp. 70–75., doi:10.1109/mic.2009.14.
14. Fujita, G., et al. “A New Motion Estimation Core Dedicated to H.263 Video Coding.” *Proceedings of 1997 IEEE International Symposium on Circuits and Systems. Circuits and Systems in the Information Age ISCAS 97*, doi:10.1109/iscas.1997.622018.
15. Greengrass, Jason, et al. “Not All Packets Are Equal, Part 2: The Impact of Network Packet Loss on Video Quality.” *IEEE Internet Computing*, vol. 13, no. 2, 2009, pp. 74–82., doi:10.1109/mic.2009.40.

16. Lancett, Peter, et al. “Peer-to-Peer vs. Client-Server Networks.” *Techwalla*, https://www.techwalla.com/articles/peer-to-peer-vs-client-server-networks.

17. “Cloud-Based and Peer-to-Peer Meetings.” *Zoom Blog*, 22 Oct. 2014, https://blog.zoom.us/wordpress/2014/10/09/cloud-based-and-peer-peer-meetings/.

18. “What Is a Protocol?” *Computer Hope*, 7 Oct. 2019, www.computerhope.com/jargon/p/protocol.htm.

19. “Transmission Control Protocol DARPA Internet Program Protocol Specification.” *IETF*, Information Sciences Insit, Sept. 1981, www.ietf.org/rfc/rfc0793.txt.

20. Shuler, Rus. “How Does the Internet Work?” *How Does the Internet Work?*, 2002, web.stanford.edu/class/msande91si/www-spr04/readings/week1/InternetWhitepaper.htm.

21. “Difference between TCP and UDP.” *The Crazy Programmer*, 8 Jan. 2018, www.thecrazyprogrammer.com/2018/01/difference-tcp-udp.html.

22. Åkesson, Oskar. “What Is the Difference between TCP and UDP?” *Holm Security*, 15 Apr. 2019, http://support.holmsecurity.com/hc/en-us/articles/212963869-What-is-the-difference-between-TCP-and-UDP-.

23. “ISO/IEC 20922:2016.” *ISO*, 8 June 2016, www.iso.org/standard/69466.html.

24. “MQTT Publish.” *Publish MQTT Messages and Subscribe to Message Topics - MATLAB & Simulink*, www.mathworks.com/help/supportpkg/raspberrypi/ref/publish-and-subscribe-to-mqtt-messages.html.

25. “Download.” *Wireshark · Go Deep.*, https://www.wireshark.org/.

26. Banerjee, Usha, et al. “Evaluation of the Capabilities of WireShark as a Tool for Intrusion Detection.” *International Journal of Computer Applications*, vol. 6, no. 7, Oct. 2010, pp. 1–5., doi:10.5120/1092-1427.

27. “Clumsy 0.2.” *Clumsy, a Utility for Simulating Broken Network for Windows Vista / Windows 7 and Above,* https://jagt.github.io/clumsy/.

28. Garcia, Ahiza, et al. “Zoom Nearly Reaches $16 Billion in Value after First Day of Trading.” CNN, Cable News Network, 18 Apr. 2019, www.cnn.com/2019/04/18/tech/zoom-ipo-video-conference/index.html.

29. Duffy, Greg. “What Is Network Latency (and How Do You Use a Latency Calculator to Calculate Throughput)?” *The Best Managed Networks and Managed IT Services You Can Buy*, 29 Apr. 2019, www.sas.co.uk/blog/what-is-network-latency-how-do-you-use-a-latency-calculator-to-calculate-throughput.

30. “Indiana University.” *What Is a Packet?*, 18 Jan. 2018, kb.iu.edu/d/anyq.

31. *MANAGEMENT INFORMATION SYSTEMS Chapter 8*, Prentice Hall, 2006, paginas.fe.up.pt/~acbrito/laudon/ch8/chpt8-2main.htm.