A Project Report

On

Benchmarking Audio-Video Conferencing Applications

BY

Dhruv Nagpal

2018A7PS0095G

Rohan Kumar

2018A7PS1013G

Yash Jain

2017A7PS0186H

Under the supervision of

Dr. Vinayak Naik

Dr. Dipanjan Chakraborty

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

CS F266/F376/367/376: STUDY/LAB/DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

GOA/HYDERABAD CAMPUS

(FEBRUARY 2021)

PROBLEM STATEMENT

During the ongoing pandemic, video conferencing apps have seen an exponential rise in usage. They have found usage in new areas like education, shopping and family events. However, most students still depend on mobile data networks for Internet access. In this project we wish to study the performance of different popular video conferencing apps, (viz. Zoom, Google Meet, Skype, Microsoft Teams) used for online teaching, in terms of networks and user experience, on different network conditions (viz. 2G, 3G, 4G, wired broadband). The different platforms will be tested on parameters such as jitter, upload and download payload, processor and memory load and bandwidth. Video Characteristics (SNR and SSIM) will be compared using tools such as msu_vqmt and Audio characteristics will be compared using spek.

Related Works

In its *White Paper on Measurement of Wireless Data Speeds, 5 February 2018*, TRAI defines Minimum Download speed, Average throughput for packet data, latency, drop rate and successful data transmission percentage as QoS parameters for LTE connections. These parameters serve as a benchmark for acceptable internet quality over mobile networks.

There are discussions on methods to objectively quantify video quality. It has been noted that while subjective human judgement is the ultimate reference standard for video quality, generic algorithms might prove to be easier to deploy in a range of situations such as AV apps, video streaming sites. In *Motion Tuned Spatio-temporal Quality Assessment of Natural Videos*, it is argued that any video quality analysis must be weighted based on spatio-temporal distortion, since video quality depends not only on resolution but also smoothness of playback. They proposed a MOVIE index for VQA. Similar attempts have also been made in other papers.

There is also relevant literature as part of workshops and seminars conducted by the International Telecommunication Union. In *Measuring the popular OTT is equivalent as measuring customer experience in mobile networks?,* there is a discussion on how existing QoS parameters need to also focus on OTT platforms since video is becoming a big part of the internet as we know it. This can be seen to culminate in how AV applications are replacing vocal or textual conferencing. The seminar discussed QoS metrics for OTT platforms, which can be useful to work with AV applications as well.

The paper, *Quality of Experience for Streaming Services: Measurements, Challenges and Insights* goes through metrics and their efficacy, viz a viz the challenges faced in using them for judging video and network quality.

Testing Methodology

Quality of Service can be judged in relative terms once suitable parameters are devised. After deciding upon some parameters, conferencing applications can then be tested and compared. It was decided that the following parameters would be compared:

1) Inter Packet Arrival Time/Jitter

2) Video Characteristics like SNR and SSIM

3) Upload and Download Payload

4) Resource Consumption

   a) Processor Load
   b) Memory Load (RAM Usage)
   c) Bandwidth

To measure bandwidth, upload and download payload, and Inter packet arrival time, we use wireshark for packet capture and then calculate these values from the obtained measurement data. Bandwidth can be calculated as:

$Bandwidth, B = \frac{\sum \text{Packet Size}}{\text{Time Interval}}$

$UploadPayload = \frac{\sum \text{Packet payload sent from source to server}}{\text{Total Time}}$

$DownloadPayload = \frac{\sum \text{Packet payload sent from server to destination}}{\text{Total Time}}$

$IPAT = median\,(Arrivaltimedifferencebetweenpacketiandpacketi + 1)$

Memory and processor load can be recorded using the top command in linux, and scripts were accordingly written to capture resource consumption by the AV conferencing application. To measure other parameters like bandwidth, upload and download payload, IPAT, we use Wireshark to capture network traffic while testing. This data can then be analysed either through Wireshark or through custom scripts to analyse the raw data and produce quantifiable outputs.

Data is collected by all parties participating in the project for more comprehensive analysis.

Source Code

https://colab.research.google.com/drive/1mothMo1IM_F0vNBAFnnBx5DBNBCy3WHp

For monitoring CPU and Memory usage on Linux systems, the following command may be used:

```
htop -b -d 1 -n 60 -p <PID> > top.txt | cat top.txt | grep <PID> | cut -c 51-61 > final.txt
```

Column 1 would be combined percentage utilization of all CPU cores, and column 2 would be percentage utilization of primary system memory.

This command uses htop instead of top because of the former's suitability for modern multicore CPUs. Top might often show CPU utilization >100% since it considers 100% as on a single core, and not the processor as a whole. htop doesn't have this drawback.

CPU utilization as shown here is the percentage of time the CPU is not idle, ie, not running the idle thread. This metric is considered acceptable for our usage instead of other metrics such as Instructions per cycle, since we are tracking application load instead of processor efficiency.

Developing Testing Frameworks

It is important to lay the groundwork for the project before data is collected. Therefore, the team has focused on developing a testing framework and ensuring that the results are working properly. This framework involves a protocol for collecting raw data, and working with tools that allow us to analyse and plot the required data.

Description of data collection protocol:

Two participants would join a call on a selected AV application, such as Zoom or Google Meet. Then, one participant would share their screen which has a video of two people playing ping pong. While the video is allowed to play for two minutes, both participants capture resource consumption of the application using top command and also capture network traffic through wireshark. Both these files are then converted into csv files to be analysed using Python Pandas and Numpy libraries. These files are available on the URL provided in the Source Code section. It is ensured that other network intensive activities such as Windows Updates, apt/aur/pacman are not in use during the test.

Description of System and Network Used

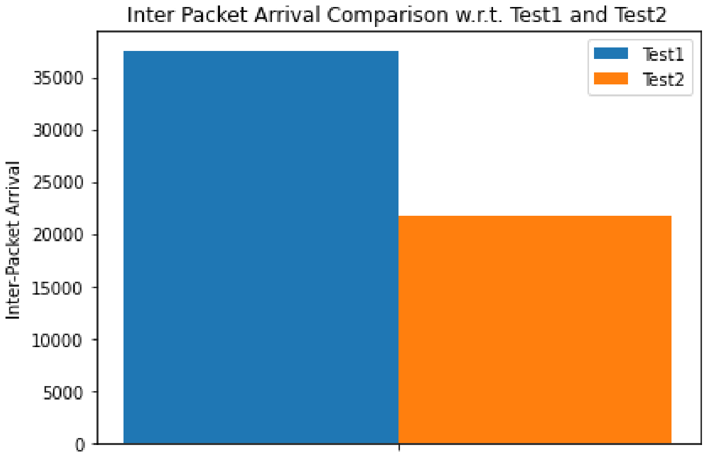
Three participants took part in the testing for the project.

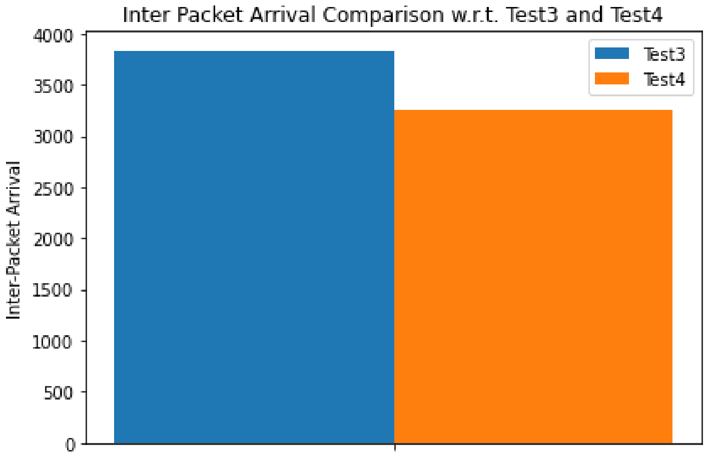
Participant #	**CPU**	**Memory (GB)**	**Operating System**	**Internet Connection Type**	**Internet Connection Speed (Mbps)**
1	i5-8265U	8 GB	Windows 10	WLAN, JioFiber	150 Mbps
2	i5-8250U	16 GB	Windows 10	WLAN, JioFiber	150 Mbps
3	i7-9750H‘	16 GB	Windows 10	WLAN, BSNL	120 Mbps

Graphs plotted and preliminary observations

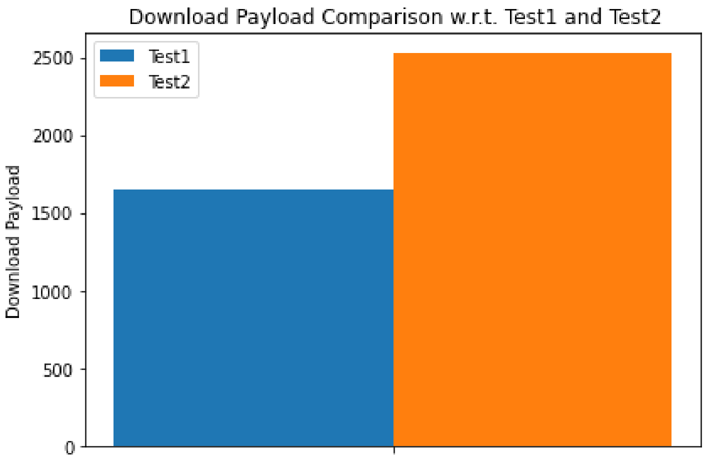
For all the plots shown below,

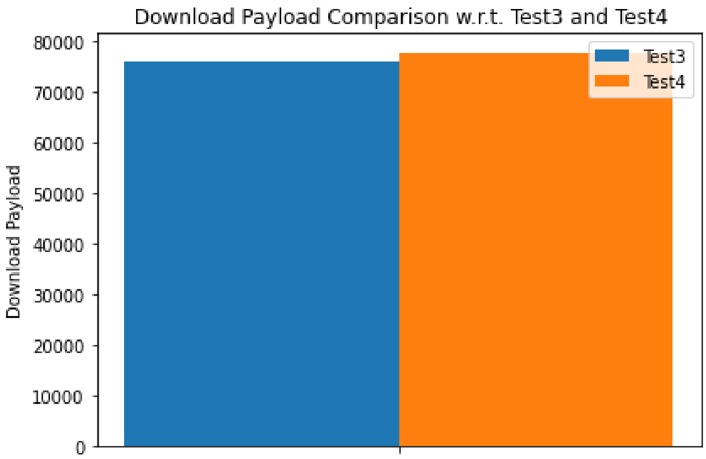
- Test 1 - No mic no cam
- Test 2 - mic on no cam
- Test 3 - mic on cam on
- Test 4 - mic on cam on with blurred background



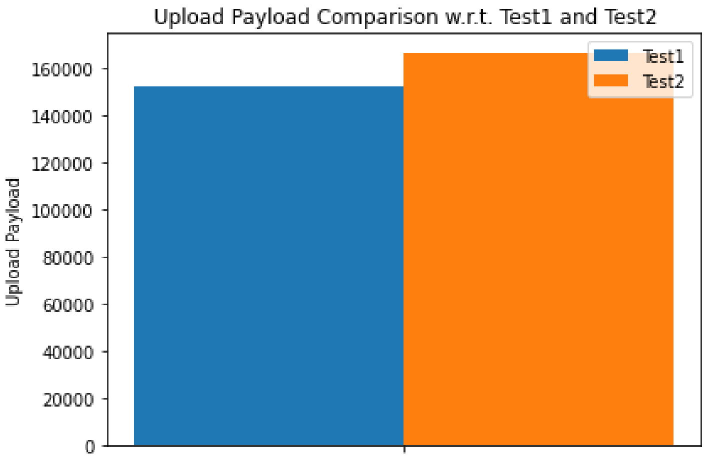


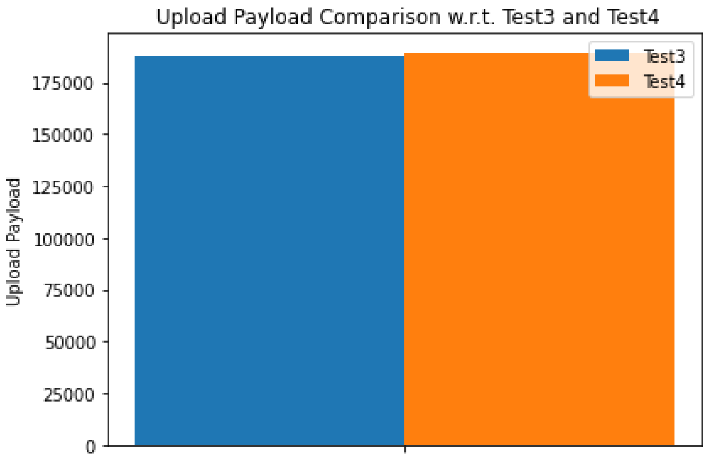
IPAT for google meet data in milliseconds





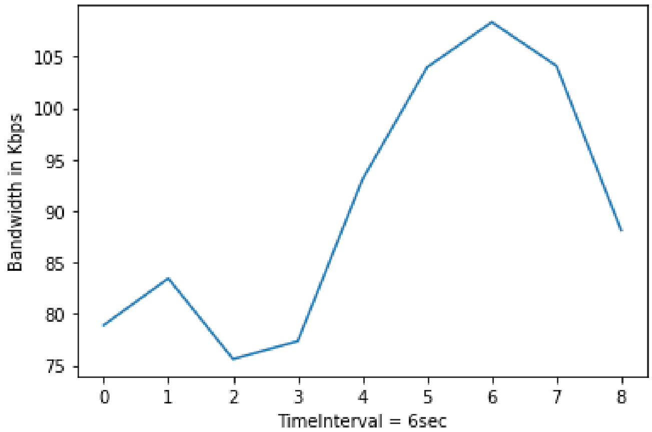
Download payload for Google Meet data in Kb





Upload payload for Google Meet data in Kb

It can be seen that the IPAT falls drastically(10x) for the tests corresponding to camera on. This is because the number of packets to be sent for video are far greater than audio. To ensure smooth video streaming, applications send out packets faster. The amount of data sent from server to destination(download payload) is very high in the video case as video data is very large compared to audio. Upload payload for video is about 10% higher than audio but the download payload is 25x higher because audio has a higher compression rate compared to video.



Bandwidth for Zoom web application(Test 1)

Bandwidth values for other applications will be plotted in the same graph for a comparative study.

Pending Work

1) Write scripts for plotting CPU and memory usage.

2) Study research papers testing CPU usage of applications to find out possible unbiased ways of plotting CPU usage for different applications.

3) Find out a plausible way to measure latency of a packet to a high level of precision.

4) Collect data from different platforms by varying network conditions. This data will be plotted and studied to conclude which applications perform efficiently under different conditions.

5) Study the protocols used by different applications and draw conclusions for their performance.

6) Using msu video quality measurement tool to see the variation in frames per second and resolution during screen sharing and using spek to compare quality of audio.

Future work

1) Similar experiments can be conducted on mobile platforms as well to see how an application performs when it has hardware constraints.

2) It is clear that any objective analysis of these networks is useful only if there is a study on performance of audio video technologies and applications on mobile networks such as LTE and upcoming 5G networks. It also involves a study of how these networks adapt to network constraints, and an analysis of their compression algorithms for reducing upload/download is needed. This analysis must be conducted through analysis of recorded video and audio, with MSU VQMT.

3) Replicating these experiments with a larger number of participants.

4) Identifying impact of inbuilt optimization options such as "optimised screen sharing" on Zoom.

Conclusions

1) The findings from this experiment can help point out possible drawbacks in application design at the network level and help improve their experience by rectifying the same.

2) Find out differences in the native application and web application from the performance point of view for the same platform.

References

Khadija Bouraqia, Essaid Sabir, Mohamed Sadik, Latif Ladid: "Quality of Experience for Streaming Services: Measurements, Challenges and Insights" , 2019; http://arxiv.org/abs/1912.11318

Seshadrinathan, K., and A.C. Bovik. "Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos." *IEEE Transactions on Image Processing*, vol. 19, no. 2, 2010, pp. 335–350., doi:10.1109/tip.2009.2034

Telecom Regulatory Authority of India, . "White Paper on Measurement of Wireless Data Speeds." , 2018

Khadija Bouraqia, Essaid Sabir, Mohamed Sadik, Latif Ladid: "Quality of Experience for Streaming Services: Measurements, Challenges and Insights" , 2019