Video Streaming Traffic Study in India

CSE 534 Final Project Report

Debasmita Basu
Stony Brook ID# 112682239

Aiswarya Lakshmi Renganathan
Stony Brook ID# 112688118

1. Abstract:

The size and complexity of the Internet continues to grow, and we know that it is absolutely loaded with video streaming traffic. Video streaming has been growing at a much faster rate than others. In India, over the last 2 to 3 years, the advent of low-cost data carriers has revolutionized the overall video content consumption. According to a popular company's newest Visual Networking Index, video will account for more than 70% of all IP traffic in India by 2022.

So, understanding video streaming metrics becomes more and more important for providers, network operators and distributors. Also, the content mostly is encrypted these days, which makes the understanding even more difficult.

2. Overview:

In general, video-on-demand services enables users to start seeing the video while the content is still being downloaded. Viewers can view the videos either on their personal computers through a web browser or on mobile devices. TCP connections are mostly used for video streaming. The entire streaming duration mainly has the initial buffering stage and the steady state.

In the last few years, video streaming services in India have grown increasingly popular. The three dominant sources for video streaming have been YouTube, Netflix, Amazon Prime. In this project, we study and analyze some of the significant metrics of the popular video streaming services traffic in India which impact the overall quality streamed and hence, the user viewing experience. We also do a comparative study of India, US video streaming traffic for these services.

3. Dataset:

For data collection, we have gathered network packets from 10 people in different parts of India. The data as of now contains network packets from three popular video streaming services. For doing a comparative study for some of the features, we also collected Wireshark traces for these apps while playing them in the US.

- The three sources for our study of video streaming traffic are:
    - YouTube,
    - Netflix and
    - Amazon Prime.

- The entire dataset has been captured over a variety of networks like Wi-Fi, 4G and 3G. The 4G and 3G data also spans around different data carriers like Reliance Jio, Airtel, Vodafone, BSNL.
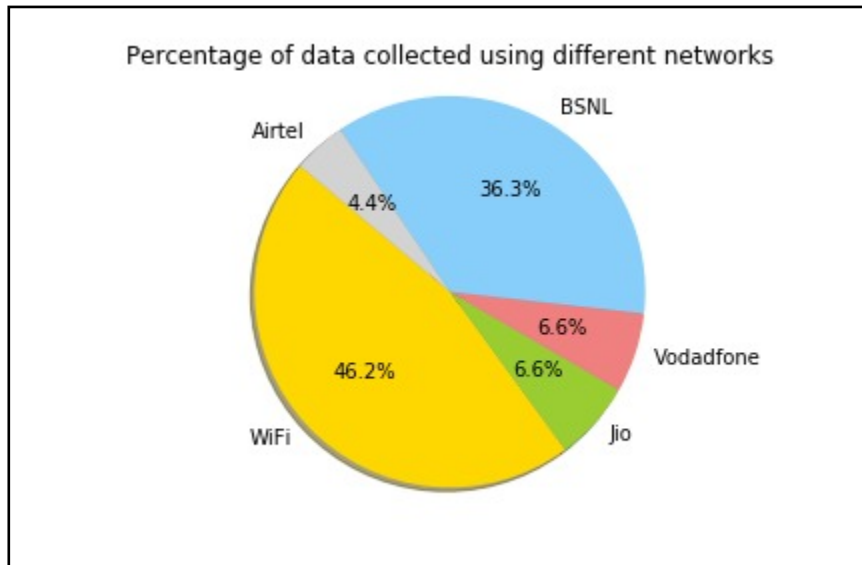

Fig 1. Percentage of data collected using different networks – India.

- For comparative study with respect to some important features, we also collect the Wireshark network traces while playing the video apps in the US over Wi-Fi, mobile networks like Verizon and AT&T.

- We also have considerable data for different video resolutions. We have divided the data into categories of low, medium and high resolutions.

Class	YouTube	Amazon Prime	Netflix
Low	<=360p	0.38GBph (Good)	0.3GBph (Low)
Medium	>360 and <720p	1.4GBph (Better)	0.7GBph (Medium)
High	>=720p	6.84GBph (Best)	3GBph for HD/7GBph for UHD (High)

4. Experimental Setup:

For the data collection phase, we have used Wireshark to capture the network packets while videos were played from the three services on web browsers. For evaluation phase also, we have done our analysis using Wireshark tool. For deriving the result, we have so far used Python module 'dpkt' to programmatically analyze the network flow, and rate of data flow.

5. Results:

I. Number of network flows in single sessions and the protocols used:

We analyzed entire YouTube, Netflix and Amazon Prime datasets spanning over different kinds, networks and video resolutions to observe the number of streams used in single sessions by different services.
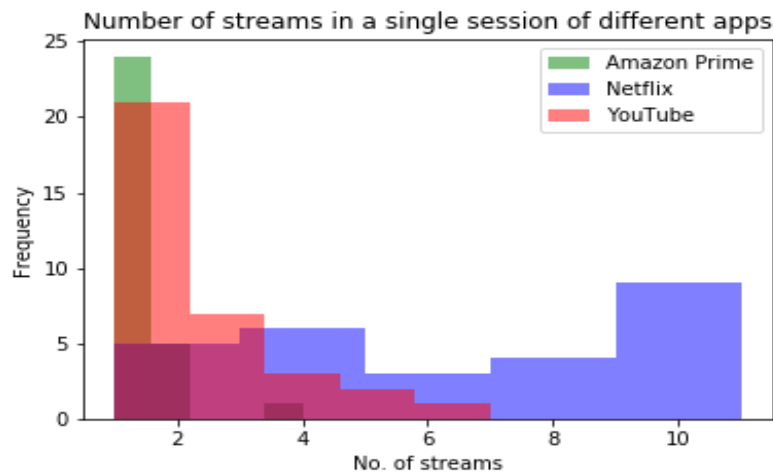

Fig 2a. Number of streams in a single session of different applications

We can observe from the above histogram plot that the number of streams used by both Amazon Prime is 1 in maximum cases. But Netflix generally uses up to 10 streams in every session. Both Amazon Prime and Netflix use TCP as the transport layer protocol. YouTube on the other hand use both UDP and TCP protocols. From our collected Wireshark traces, we find that with YouTube UDP is more popular with 62.5% sessions using UDP and only 37.5% using TCP, as shown in Fig 2b. The ones that use TCP, have close to 6 streams per YouTube session.

This information serves as a good checkpoint for us to further investigate the different application layer protocol types used by these applications.

Since Amazon Prime is running on AWS, HTTP/2 is the protocol used by it. YouTube UDP uses QUIC.

Netflix uses HTTP 1.1 and not HTTP/2. After investigation, it has been found out that Netflix has identified several denial of service (DoS) flaws in numerous implementations of HTTP/2. Netflix

quoted "The algorithms and mechanisms for detecting and mitigating "abnormal" behavior are significantly vague and left as an exercise for the implementer. From a review of various software packages, it appears that this has led to a variety of implementations with a variety of good ideas, but also some weaknesses.".
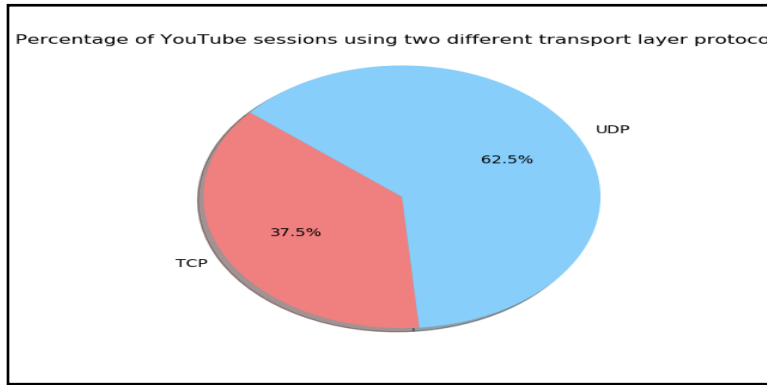

Fig 3b. Percentage of YouTube sessions using UDP and TCP.

II. RTT:

Round-Trip Time (RTT) of the three applications while playing in India, have been compared in Fig 3 below. Three websites of the video services were pinged from India and the unloaded RTTs were plotted.
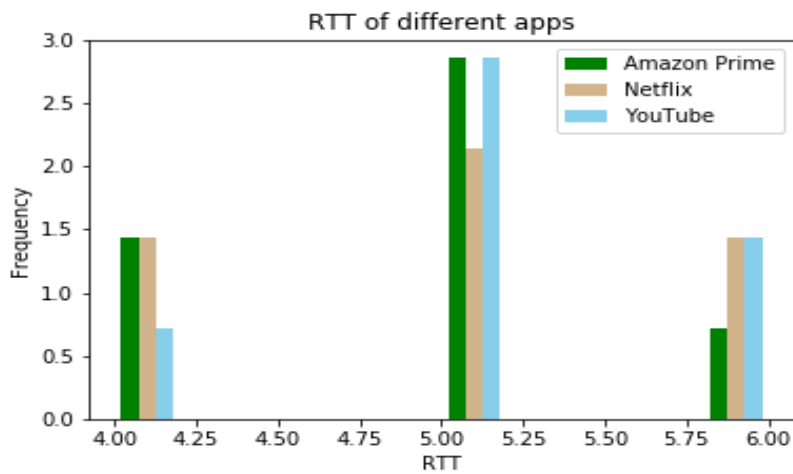

Fig 4. RTT distribution of different applications compared

From the above graph, we can see that the RTT of all the three applications (played in India) are similar and distributed around 4 to 6 ms.

III. Throughput:

We use the following throughput traces to make a comparative study on datasets from both India and US:

- YouTube dataset contains bits per sec throughput measurements from web browsers.
- Netflix dataset contains bits per sec throughput measurements from web browsers.
- Amazon Prime dataset contains bits per sec throughput measurements from web browsers

From the Wireshark traces gathered, we calculate the individual throughput in bits per second of all the samples and plot a CDF Fig 4. shows the CDF of the average throughput in India and US for all the three video services. Fig 5. and Fig 6. show the average throughput across Wi-Fi and different mobile networks respectively, in both India and US for the same apps.

Between the video streaming services, most data is consumed in the following order, in both the countries:

1.YouTube
2. Netflix
3. Amazon Prime

When comparing between different apps and the country where they are streamed, we find the data consumption rates to be roughly in the following order (Fig 4.):

1. YouTube-US
2. YouTube-India
3. Netflix -US
4. Netflix-India
5. Amazon Prime-US
6. Amazon Prime-India

Hence while YouTube is the most data consuming video application in both the countries, all the three video streaming services have higher throughput in the US as compared to India.
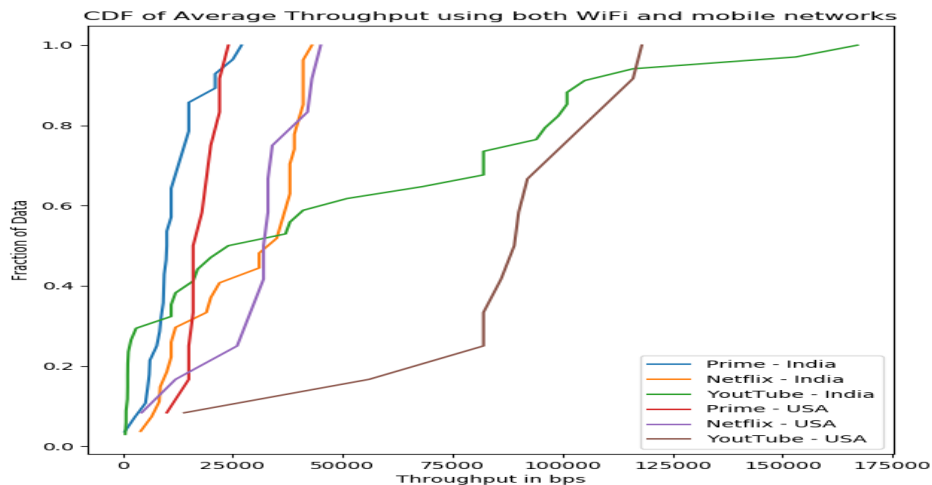

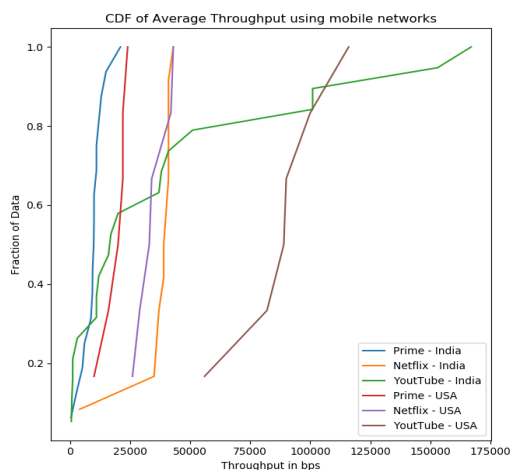*Fig4. Throughput comparison of video apps streamed in India and the US.*

Fig 5. Throughput of video apps streamed via mobile networks – India vs US
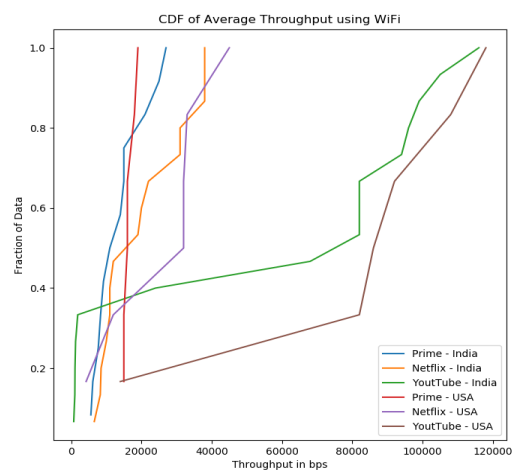

Fig 6. Throughput of video apps streamed via Wi-fi networks – India vs US

We also plotted CDFs to see if the throughput appears to be different while considering only Wi-fi or mobile networks. YouTube still exhibits the maximum throughput, followed by Netflix and Prime over both Wi-Fi and mobile networks. But interestingly, over Wi-fi, data consumption by Netflix-India comes to be more than Netflix-US.

The throughput of the three video streaming apps in US, appear in the same order as in India, be it across all networks or just Wi-Fi / mobile network.

Bandwidth rates vary widely depending on the media format and quality. From the above graph, we observe that the Throughput of YouTube is significantly greater than both Netflix and Amazon Prime.

Of all the three applications, Amazon Prime has been found to be having the least Throughput from our initial study. But as we know, Amazon uses HTTP/2 whereas Netflix runs on HTTP/1.1. Yet, the throughput of Netflix is found to be comparatively higher than that of Amazon. This is because of the use of Open Connect by Netflix.

"Netflix CDN" team was launched in the middle of 2011. NETFLIX's CDN is called Open Connect. Netflix has deployed its very own Content Delivery Network (CDN) to mainly address single point of failures, latency issues in locations away from the origin servers. NETFLIX works directly with the ISPs and installs their own boxes called Open Connect Appliances at either Exchange Points or in the ISP facilities themselves, holding up to 280 TB of video data each (almost the entire NETFLIX library). Now that NETFLIX has its own ISP, the NETFLIX data packets don't have to fight with all the other internet traffic that is upstream from the ISP. NETFLIX's appliances push out data at over 90GB/sec=13,000 people watching NETFLIX at once. Cataloguing is done in the morning when there is less traffic.

IV. Burst Analysis:

Video traces as observed from Wireshark, can be divided into three phases:

- On-time: Time during which data is transferred, when client requests for it.
- Off-time: Time during which there is no data transfer.
- Bursts: The arbitrary areas of data intensity in the traces are known as bursts.

From our data, we observed the traces for the initial 30 secs to find which application is more bursty. We can see from the following figure, that the number of bursts per 5 seconds in Amazon Prime is much higher while YouTube has the least number of bursts. Also, the burst size (no of packets) per burst is the highest for Amazon Prime.
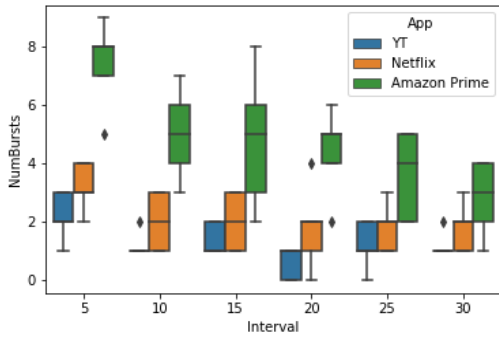
We find this true for both India and US.

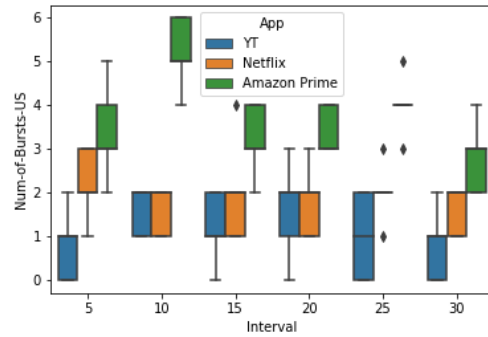Fig 6. No of Bursts per 5 secs - India

Fig 7. No. of Bursts per 5 secs - US

We also plot the sizes of the bursts (no. packets in each burst) with time, for the first 30 secs of playing the applications. In Amazon Prime, which according to the traces is more bursty, also happen to inject more packets into the stream with each burst. YouTube's burst sizes remain more or less uniform with time.
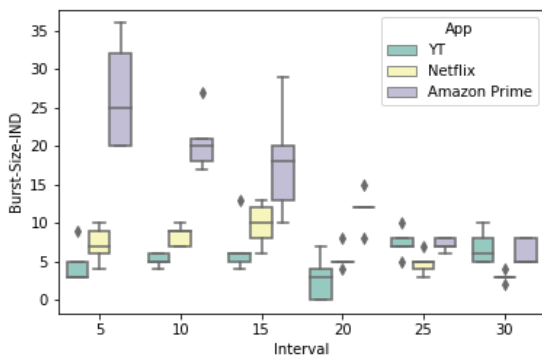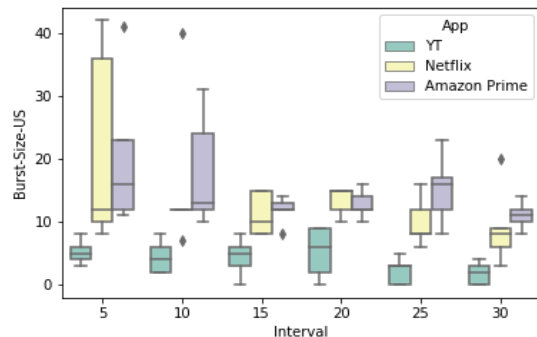

Fig 8. Burst sizes with time - India

Fig 9. Burst sizes with time - USA

The number of bursts and their sizes happen to directly affect the video streaming performances. Quality of Service (QoS) is the collective effect of service performances, which determine the degree of satisfaction by a user.

Video streaming requires packets to be transported by the network from the server to the client. As long as the network is not too heavily loaded the stream will obtain the network resources (bandwidth) that it requires. However, once a certain network load is reached the stream will experience delays to packets and packet loss. In real-time video streaming, often requiring the transfer of large amounts of data packets across a network, any loss of data directly affects the seamlessness of the video stream at the receiver. So, the frequency of bursts and their sizes have a role in the users' overall perception of the QoS.

If traffic is buffered at the receiver it may be possible for the re-transmission of missing data to be requested. So, when huge bursts occur, huge amounts packet loss occurs in the network. A snapshot of Wireshark I/O of a Netflix stream helps to visualize this better. We see that the bursts of data, trigger almost simultaneous losses (black bars in the data peaks).
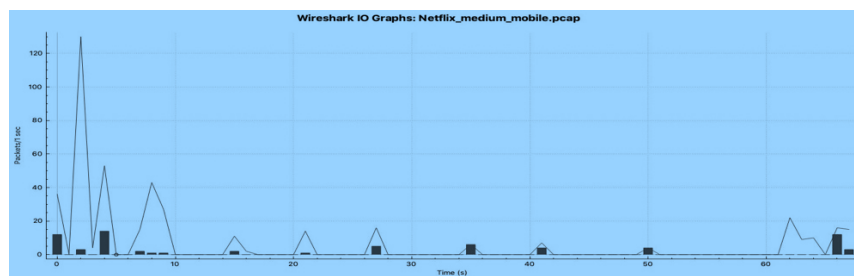

Fig 10. Wireshark I/O Graph for Netflix showing bursts and losses.

6. Discussion and Future Work:

This study can be generalized with more samples. With more observations our claims will have stronger empirical support. The current study is the mean behavior of streaming services across multiple network, devices or browsers in a specific country. It'll be interesting to observe the app behaviors with different combinations of browsers, devices, with fixed network configurations. In this way, it'll be possible to understand the most dominant network component of a specific streaming service. We'll be able to answer the question, how our observations depend on the app or the browser or the device.

To further extend the study we can look into the following questions:

1. What are the encoding techniques used by the different streaming services?
2. Are Amazon Prime/YouTube affected by HTTP/2 DoS flaws identified by Netflix?

7. Conclusion:

Since video accounts for majority of the network traffic, understanding video streaming metrics is an important aspect for providers, network operators and distributors. In this project we study some behaviors of the three top video streaming services – YouTube, Netflix, Amazon Prime. We see from the data gathered for this study that YouTube has the maximum throughput, Amazon Prime

is most bursty while streaming. Also, from the number of flows used by the apps in a single session, we identify the application layer protocols popularly used by them.

8. References:

1. Rao, Ashwin, Arnaud Legout, Yeon-sup Lim, Don Towsley, Chadi Barakat, and Walid Dabbous. "Network characteristics of video streaming traffic." In Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies, p. 25. ACM, 2011.
2. Mangla, Tarun, Emir Halepovic, Mostafa Ammar, and Ellen Zegura. "eMIMIC: estimating http-based video QoE metrics from encrypted network traffic." In 2018 Network Traffic Measurement and Analysis Conference (TMA), pp. 1-8. IEEE, 2018.
3. Ravattu, Radha, and Prudhviraj Balasetty. "Characterization of YouTube Video Streaming Traffic." (2013).
4. https://nakedsecurity.sophos.com/2019/08/19/netflix-finds-multiple-http2-dos-flaws/
5. https://ma.ttias.be/googles-quic-protocol-moving-web-tcp-udp/

9. Project repository:

https://github.com/aiswaryalakshmi/CSE534-FCN-Project