Live vs. Video on Demand within a VPN Detection.

Andrey Pristinsky[1] Da Gong[2] Mariam Qader[3] Tianran Qiu[4] Zishun Jin[5]
[1]Halıcıoğlu Data Science Institute,
University of California, San Diego

I. INTRODUCTION

Due to the variety, affordability and convenience of online video streaming, there are more subscribers than ever to video streaming platforms. Moreover, the decreased operation of non-essential businesses and increase in the number of people working from home in this past year has further compounded this effect. More people are streaming live lectures, sports, news, and video calls via the internet at home today than we have ever seen before. Internet Service Providers, such as Viasat, are tasked with optimizing internet connections and tailoring their allocation of resources to fit each unique customer's needs. With this increase in internet activity, it would be especially beneficial for Viasat to understand what issues arise when customers stream various forms of video.

In general, different internet activities require different resources to optimize the connection. For example, if a customer watches a lot of live video they may prefer a connection with lower latency. A live stream requires low latency in order for the streamer and audience to communicate in real time without a significant lag. As latency increases, the delay in time between the audience receiving the video from the streamer (lag) increases. Live streaming usage of latency is dependent on what quality is acceptable. For instance, any latency larger than nearly real time could be troublesome in zoom calls where one wants to be able to have conversations between others fluidly, while low latency that is higher than zoom latency is acceptable in streams that have slight to no interactivity, where a slight delay of 10 or 20 seconds wouldn't't affect the end user, such as a news stream.

Although we are able to identify the genre of an activity when a user is not using a VPN, the challenge arises when a user chooses to surf the web through a VPN. When it comes to VPN use cases we can't identify a user's unique activity when they experience issues, thus making us unable to successfully troubleshoot those problems. This is where a tool that could identify various internet activities, specifically live or uploaded video streaming, within a VPN tunnel would be extremely useful for an Internet Service Provider.

In the past, there have been effective ways to distinguish video streaming from general internet activity. Yet, distinguishing between different types of video is a little more challenging. User's experience both live video and video on demand in the same way over the internet; users can play videos while the video platforms send the proceeding content making the video stream smooth with minor buffers. However, live video has a few components that video on demand does not. Some live videos have an interactive component where the audience can communicate with the streamer in real time. Live streams also do not have quality controls, where users can set the quality of the video to a certain level. Our goal is to distinguish how providers send live video vs. pre-recorded videos to their users. Other works have successfully achieved classifying the two types of video by looking at the payloads of packets, however their methods do not work with encrypted data since we are unable to see packet's raw data [2].

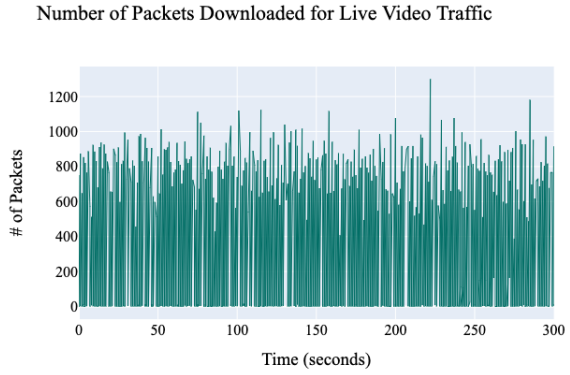
Although we cannot see the contents of the packets transported across a network, patterns in the way they are sent can still help us identify if a user is streaming a live or pre-uploaded video. Our task will be to detect key differences between the way information is sent across a network for live streaming and video on demand. To achieve this, we will generate an extensive data set consisting of network data for both types of video streaming. To create the data we will use a tool that connects to our own interfaces and captures consistent real time internet traffic from our personal devices. Initially, we will focus our efforts on a platform that offers both types of content, such as Twitch. By capturing several chunks of video data from Twitch using a VPN, we will create a large data set that can help us understand the patterns video streaming providers use to deliver their content. Eventually, we hope to broaden our scope to multiple platforms and content types (i.e. live video streams vs conference

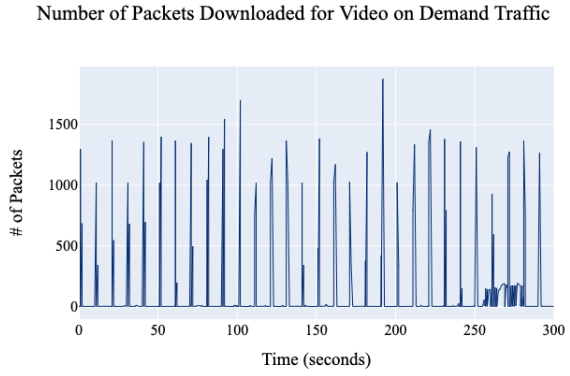
calls).

II. METHODS

The internet data that we have collected consists of the number of packets and bytes being uploaded and downloaded across a connection. A connection consists of the source and destination IP addresses and ports.With this information, we can potentially find significant features that are key identifiers of internet activity. Using this data, we can look at the flow of packets and bytes sent back and forth over time between the user and destination. Through these findings, we plan to create a machine learning model to predict if a user is streaming live or pre-uploaded video.

Similar to other common approaches to analyze internet network data, we have chosen to look for statistical differences between the flow of packets across a network for live video streaming and vod. The graphs below look at the number of packets sent across a network over time for both twitch live and twitch uploaded videos.





When looking at the graphs above a few differences are immediately apparent. First, we can see that the live video has a denser graph with more packets coming in more frequently. On the other hand, the vod has more time between each spike but the magnitude of packets coming in at a time is larger. To quantify this key difference, we can take the ratio of time packets are being sent to the time packets are not being sent (packet size is 0). This will tell us how much time during the viewing of the video no packets were being sent from the destination to the user.

However, there are many micro spikes that couldn't be observed from the graphs, which would affect the accuracy of the previous method. For example, a noisy VOD traffic may have many small size packet transactions resulting in the ratio of packets transferring being as high as the live video streaming. To eliminate this possible error, we calculate the time between each spike as leisure time (the gap in seconds between each spike). Typically, VOD has more leisure time and live streaming has less. Live streaming requires video providers to consistently send data to their users as they are sending it in real time, this is a key difference in the way live streaming vs. VOD is delivered to viewers.

Another way to quantify the difference in density of the two video streams is by simply looking at the number of peaks present. There are considerably more spikes in the dense graph for live streaming at smaller sizes, compared to the more spaced out larger spikes in the VOD plot. With the features we have extracted to distinguish live streaming from VOD, we can use a machine learning model to predict the type of video a chunk of internet data is.

III. MODEL

Since we are predicting a binary result of whether the file is VOD or live streaming, we explored classifiers including the SVM, KNeighbors classifier, Logistic Regression classifier and Random Forest classifier. Random Forest classifier achieved the highest accuracy of 99 percent, which is nearly 15 percent better than other models. The possible reason why Random Forest Classifier has the highest accuracy is that the more features we train it, the higher the accuracy would be. However, this classifier takes on average five times longer than other three classifiers. Eventually we trained the model on Zoom data and Twitch traffic and achieved the 99 percent test accuracy. Moreover, we noticed that if we included the YouTube data in the training process, the accuracy of the overall model would be lower by 10 percent. This may due to the fact that YouTube used a different algorithm than zoom and twitch which messed up the model.We plan to explore youtube data further in the coming weeks, and hyper tune our classifier to be able to distinguish different live

video content. We also hope to explore the differences between live and VOD within the frequency domain to gather more insight.

IV. Results

A. Trained Binary Classifier

B. Predictions on Test Set

V. Discussion

A. Accuracy of the Classifier

B. Details about Network Data

VI. References

1) https://www.boxcast.com/blog/live-stream-video-latency

2) https://www.idc.ac.il/en/schools/cs/research/documents/online%20classification%20of%20vod%202013.pdf

VII. Appendix

A. Proposal

Due to the variety, affordability and convenience of online video streaming, there are more subscribers than ever to video streaming platforms. Moreover, the decreased operation of non-essential businesses and increase in the number of people working from home in this past year has further compounded this effect. More people are streaming live lectures, sports, news, and video calls via the internet at home today than we have ever seen before. Internet Service Providers, such as Viasat, are tasked with optimizing internet connections and tailoring their allocation of resources to fit each unique customer's needs. With this increase in internet activity, it would be especially beneficial for Viasat to understand what issues arise when customers stream various forms of video. In general, different internet activities require different resources to optimize the connection. For example, if a customer watches a lot of live video they may prefer a connection with lower latency and higher bandwidth. Although we are able to identify the genre of an activity when a user is not using a VPN, the challenge arises when a user chooses to surf the web through a VPN. When it comes to VPN use cases we can't identify a user's unique activity when they experience issues, thus making us unable to successfully troubleshoot those problems. This is where a tool that could identify various internet activities, specifically live or uploaded video streaming, within a VPN tunnel would be extremely useful for an Internet Service Provider.

In this past quarter, we created a binary classifier to identify if video streaming took place inside a VPN tunnel. Similarly, next quarter our group plans to create a classifier that can predict if a user is streaming live or uploaded video. Using the same techniques we acquired from quarter one, we plan to explore how data is uploaded and downloaded across the network, as well as exploring new approaches such as analyzing internet protocol behavior.

We have used the tools provided by Viasat, including network-stats, to collect internet traffic data from students in our class. Each student collected network data while watching video and browsing the internet which we all shared with each other in class. For our project next quarter, our team will use this same method from the previous quarter to collect and share data within the group. Regarding the information we need for the project, they are included in the traffic data we collect. Since our project is about differentiating VPN encrypted live streaming videos and video on demand, we need the information about the data being transferred within the VPN tunnel, between user and server, every second.

This traffic data, and related information we need for the project, is recorded by the network-stats tool in csv files. Within the csv files, there is comprehensive information about the internet traffic that will help us analyze the data. We can extract useful features out of the data like the average, median packet size and spike numbers. These features can effectively distinguish video on demand streaming and live streaming. For example, with the collected data and its visualization below, we can quickly see the difference between the two types of streaming. (Upper is VOD, lower is live streaming). Therefore, the data is at sufficient quality for distinguishing the VOD and live streaming activity.

The project output will be a pipeline on GitHub indicating its functions. The users who want to know if there is video streaming in the dataset can simply pull the repo and run the pipeline. There will be two classifiers in the pipeline. The first classifier will determine whether the input data file was collected under live streaming video (including video conference) or video on demand. If it is classified as live streaming video data, it will then enter the second classifier. The second classifier will determine the live streaming video category of the input data. It will determine if it is live streaming game video or live streaming music video with static background or video conference call. Each classifier will return the accuracy along

with the confusion matrix to show performance. The project will also include a report. The report will show the steps or methods involved in the pipeline along with visualization of the data. It will also include an explanation of the result. The report will provide a clearer overview of the project.