DETECTING LATENCY ANOMALIES

Ridwan Chowdhury, Andrew Riordan, Shatayu Kulkarni

PURDUE UNIVERSITY

INTRODUCTION

What is Viasat? What do they do?
Viasat is a communications company based in California, although they have operations across the United States and in other countries as well. They provide high-speed satellite broadband services and secure networking systems for both military and commercial groups.

What is our project?
As part of the Purdue-Viasat Corporate Partners team, our job for the past 9 months was based on analyzing large sets of ping-latency data in order to find anomalous patterns. In doing so, our main goal was to identify various issues that could be affecting the internet service provided to Viasat's customers. In particular, we wanted to analyze the latency data for internet service on airlines using Viasat's internet service.

What is ping latency?
A ping is a signal that's sent from one computer to the other on the same network. Latency is the time it takes for the ping to make a round trip from one computer to another and back. Ping latency is a very important measurement for analyzing both the speed and reliability of a network connection. High latency is the culprit behind things like your Netflix/YouTube videos buffering every few seconds, or your video game lagging. It's important to have low latency and secure pings, especially for large commercial companies and the government.

How are we going to accomplish our goal?
There was quite a large amount of data to analyze overall, about 50 GB of ping data, and we each had ideas regarding how to go about it. We decided to go with a three-pronged approach to analyzing the data in order to obtain the most conclusive results, and have outlined each of our three approaches in detail throughout this poster.

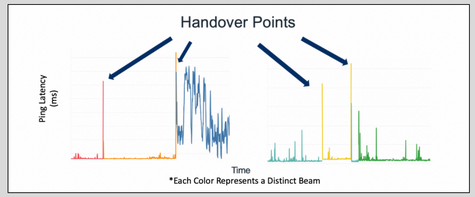
What technologies did we use?
Each individual approach used more specific technologies, but overall we utilized Python3, the Pandas and Plotly modules, and Jupyter Notebook to implement and organize our code.

APPROACH 1: INVESTIGATING HANDOVER POINTS

- One important element of the data we were given was the "beam" in which the ping took place.
- Each flight is served by a variety of these beams along its flightpath. Each beam can be visualized as a radius inside of which the flight is served by the given beam. Additionally, each beam "belonged" to one of four satellites in the data set.
- When looking at the time series data for ping latency on a variety of flights, it became clear that the "handover" points, those where the flight's service beam was changing, seemed to induce spikes in the graph. These spikes were found to be significant in the flight metadata and to last a varied amount of time.
- To look at variation amongst spikes, with respect to length and size, we defined what we meant by "spike" and a variety of other related details and scraped through all available data to compile this information about handover induced spikes in isolation.
  In addition to the analysis at right concerning satellites, we were also able to isolate the beam transition points which seemed to perform best and worst. Many of our results were consistent with internal theories about handover severity.

FIGURE 1

A graph of ping data across multiple beams. Observe ping spikes at handoff points.



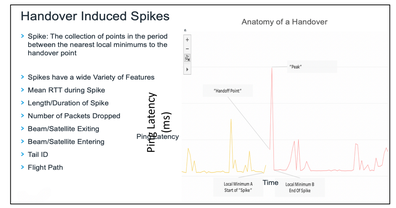


FIGURE 2

An example of a handover induced spike with labeled features.

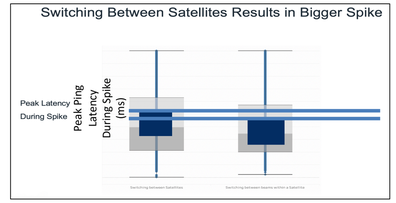


FIGURE 3

A boxplot demonstrating reduced latency spikes when beams shared a satellite.

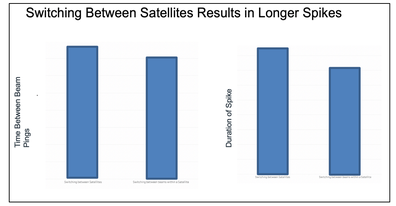
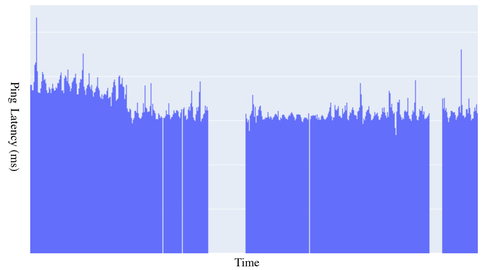


FIGURE 4

Bar charts demonstrating reduced spike times for transitions when beams shared a satellite.

FIGURE 5

A time series of the average ping latency over time for a standard satellite beam



APPROACH 2: TIME SERIES ANALYSIS

What is time series analysis?
Time series analysis is a statistical analysis technique that deals with time series data, i.e. data that occurs over a set period of time. It allows us to identify trends in data that are inherently tied to time. In Python, the Statsmodel module allows us to perform simple time series analysis on a dataset.

How can we use this to look at ping latency?
Because latency refers to the round trip time of a ping, ping latency is inherently tied to time. Using time series analysis has a high chance of producing something statistically significant.

How can we manipulate the data to analyze patterns? – Figure 5
Through the use of python and bash scripting, we can divide the data by satellite beam, and analyze the pings to a satellite beam over a longer period of time. The histogram above in Figure 5 displays the ping latency of a normal satellite beam over time.

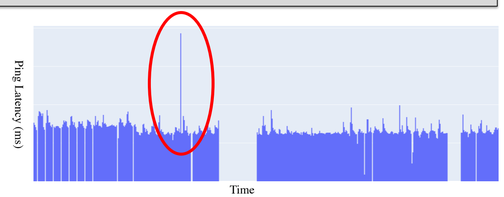
TIME SERIES: PATTERNS IN THE DATA

Overall, there are 2 major patterns in the data that we see after utilizing time series analysis.

1. **Major latency spikes – Figure 6**
   Across many beams, there are times where ping latency spikes drastically. The times during which these beams spike is consistent across many beams. These indicate large scale events that cause major ping increases to all beams, such as handovers.

FIGURE 6

A time series of the average ping latency over time for a spiking satellite beam



2. **Baseline Ping Shift – Figure** 3
   The average ping level shifts at a certain time period. This is consistent across many beams, and is indicative of some major event or change occurring to all beams. Identification of this event helps future maintenance of the beams.

FIGURE 7

A time series of the average ping latency over time for a baseline shift beam

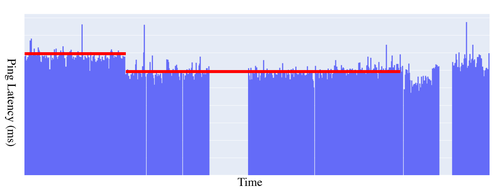


FIGURE 8

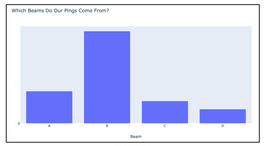


FIGURE 9

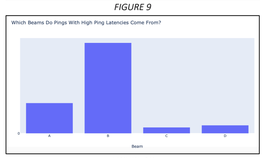


FIGURE 10

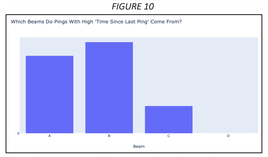
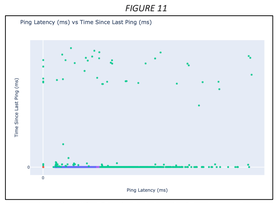


Figure 9 appears very similar to the figure 8, meaning that each beam is equally likely to give off late pings. However, Beam A's bar in Figure 10 is much bigger than its bar in Figure 8, suggesting that it should be investigated further.

FIGURE 11



A scatterplot of the studied pings (outliers in green)

APPROACH 3: USING ANOMALY DETECTION ALGORITHMS

Another approach we took to finding anomalous pings was plotting each ping's latency against how much time passed since the arrival of the ping before it and then employing an anomaly detection algorithm called DBSCAN to identify outlier points. We studied a random sample of 10% of the pings proportional to the total amount of data.

The motivation behind using DBSCAN was straightforward; the algorithm automatically detects clusters of data points and classifies points that do not fit neatly into clusters as outliers. By employing an algorithm to classify outliers, we could create a singular, consistent standard to apply to every beam studied. We could also easily extend this analysis to even more beams without having to manually tweak parameters.

Our algorithm discovered two kinds of outliers (pictured in green) - those with high ping latencies and those with high "Time Since Last Pings"; in other words, pings which arrived unusually late. It also told us that the two had essentially zero correlation. However, studying each kind of outlier separately allowed us to identify the source of these unusual pings and make analyses accordingly. We found that the studied beams are equally likely to give out pings with high latencies, but that beam A had an unusually high chance of giving out a late ping when compared to beams B, C, and D.

CONCLUSION

Through our three approaches, we discovered that unknown events on certain dates degrade Viasat's network performance, that switching between satellites strains service, and that some beams are prone to giving out both extremely late pings and pings with high latency. Our three-pronged approach proved to be effective since each separate approach found different anomalies in different datasets. In the future, we plan to extend our analyses to more of Viasat's database and also incorporate additional beam and flight data provided by the Mobility and Analytics team to find and explain more anomalies in Viasat's data.

ACKNOWLEDGEMENTS

We would like to thank our Viasat mentors, Mark Rapo and Gideon Shalev, for the help and guidance they've provided us throughout the year on our project.