Network-Topology Project

Rahul Ramachandran
cs21btech11049

Rishit D
cs21btech11053

Suryaansh Jain
cs21btech11057

August 25, 2023

Contents

1 Deliverables **1**
1.1 Introduction 1
1.2 Data 1
1.3 Visualization 1

2 Findings **2**
2.1 General 2
2.2 Latency 2
2.3 Degrees 2

1 Deliverables

1.1 Introduction

The project aims to visualise the topology of the internet. For this purpose, 4 python scripts were used: `gen_data.py`, `parse_data.py`, `graph_data.py` and `graph_gen.py`. The first script, `gen_data.py`, runs the traceroute command on a list of websites to obtain information about the path taken by the packets. This information is parsed by `parse_data.py`, which uses `whois` to perform an ASN Lookup and obtain ASNs, organization names and IP ranges. Finally, `graph_data.py` and `graph_gen.py` use this data to generate interactive graphs of the Network Topology.

1.2 Data

The raw data was obtained by running the traceroute command on 50 websites. The websites chosen include popular websites like `google.com`, and several university websites from diverse locations. The data from the traceroutes is present in `data/<foldername>/output`. This contains the latency values (in ms), the domain names and the IP addresses in the path. This data was parsed by `parse_data.py` to obtain json objects containing the ASN, organization name and IP ranges. The latencies were converted to latencies between hops, and the jsons were stored in `data/<foldername>/json`. The data was then converted to graph nodes and edges using `graph_data.py`, and stored in `<foldername>/graph_json`. Finally, a select number of destinations was chosen, and the jsons from 7 different sources were combined to form interactive graphs and a histogram.

1.3 Visualization

The results are present in the folder `visualization`. Using `pyvis`, we generated interactive graphs of the network topology. In `graph_normal.html`, the edges are colored according to the packet des-

tination. In `graph_latency.html`, the edges are assigned a color based on the latency of the route: warmer colors indicate higher latency. The third graph is a histogram of the degrees of the nodes: nodes with a higher degree correspond to exchange points and large ISPs. 3 icons are used to represent the nodes: a globe for websites, a server for the intermediate nodes and a laptop for the sources. The graphs have 7 sources and 19 destinations.

2 Findings

2.1 General

- In the first visualization, we observe that some servers are dedicated to hosting information exclusively, whereas other servers seem to host data that is not always associated with them. Such an example can seen while tracing the paths of `primevideo.com` and `ibibo.com` which both end up at the node `AMAZON-02, US` (AS Number: 16509). This is common among big-tech companies like Amazon, Microsoft, Google who host servers for websites not associated with their products.

- Over the raw-data parsed for the various visualizations, we have observed cases where the same AS Number can have multiple ranges of IP prefixes. For instance, one traceroute to `bbc.com` had two jumps both corresponding to the AS Number 55836 (`RELIANCEJIO-IN Reliance Jio Infocomm Limited, IN`) but two distinct IP addresses : 115.247.100.29 (in the range 115.240.0.0/13) and 49.44.220.181 (in the range 49.44.128.0/17).

- We observe that most hops are geographically sound. For instance, requests to `uio.no` (based in Norway) and `su.se` (based in Sweden) first reach the node `NORDUNET, DK` (Denmark, AS Number: 2603) before diverging to each of their dedicated servers, which can accounted for due to close proximity between the Nordic nations.

2.2 Latency

1. Many interesting observations can be made from the latency graph. For instance, IIT Hyderabad's traffic is routed to the NKN or Reliance JIO. The NKN is highly congested, while the latency values are lower for JIO.

2. The hops between two ASNs which are geographically distant usually have a higher latency. For instance, the hops from TATA to AS6453 in the US, from Airtel to an ASN in South Africa, from US to Australia etc. all have have high latency values.

3. A VPN was used to obtain the traceroute information for `data_Poland` and `data_US`. The initial hops have a high latency, as they likely correspond to the time taken to tunnel to the VPN servers.

2.3 Degrees

1. Nodes with a high degree are usually exchange points or large ISPs, as the degree is indicative of incoming traffic.

2. Cloudflare's autonomous system (ASN13335) has the highest degree. This is because Cloudflare is a CDN, and it receives a high amount of traffic. Other high indegrees in the US include large ISPs and backbones like Cellco Partnership (AS6167), NTT, Cogent Communications, etc.

3. In India, large regional ISPs like Reliance Jio (ASN55836), Tata Communications (ASN4775) and Airtel (ASN24560) have the highest degrees. Another thing to note is that the ASs associated with these organisations might cover a wider range of IP Addresses, and might therefore swallow up a lot of the incoming traffic.