AI3020 Assignment 1 Report

Chintalapudi Abhiroop - AI20BTECH11005
Manikanta Vallepu - AI20BTECH11014
Nelakuditi Rahul Naga - AI20BTECH11029

Raw Data Collection

We collected raw traceroute data from each of 5 sources to 10 destinations. Initially, we extracted data for 20 different destinations, but chose 10 destinations out of them to keep the visualization compact. The sources that we have used are as follows :

1. Laptop (Off Campus using Wifi)
2. Laptop (On Campus)
3. Mobile Hotspot
4. Looking Glass - London
5. Looking Glass - Tokyo

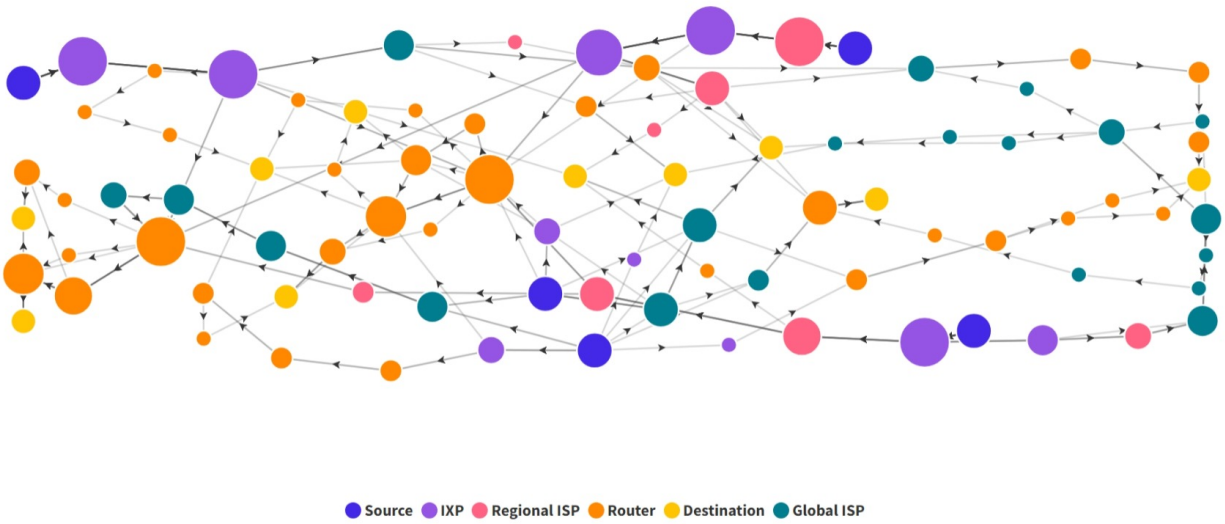
We examined various destinations such as **Google**, **Facebook**, Microsoft, **BBC**, NY Times, Flipkart, **Wikipedia**, Stackoverflow, Ebay, **Spotify**, **Twitter**, CNN, Amazon, The Guardian, **WhatsApp**, **Yahoo**, Uber, **YouTube**, Zoom and **Dropbox**. The destinations which are **bold** are the ones we have finally chosen for our visualization. They were chosen so as to maintain the diversity in the destinations. For the first 3 sources, traceroute was performed using the **tracert** [**destination_address**] command in the Windows terminal. For the Looking Glass sources, the traceroute was performed using Hurricane Electric Looking Glass website. Additionally, we also collected information such as IP address range for different organizations (according to their AS numbers) and formulated the information in the file named **ASN-IP-RANGE.csv**.

Data Preprocessing

We used Flourish Studio to visualize the Internet Topology as a directed graph. We needed to supply the nodes and edges (csv) files to the software to create a visualization. We initially tried to write Python code to create the csv files from the raw data files but faced difficulties mainly due to the fact that many IP addresses map to the same AS number and organization. Also, node type (such as Source, IXP, Router, Global and Regional ISP, Destination) was difficult to decipher using code. Hence, in the interest of time we decided to manually create the nodes and edges files (from the raw data files) named **Points.csv** and **Links.csv** respectively. Additionally, we have also submitted code (in a file named **Points_Degree.ipynb**) that calculates the total degree of each node (which is proportional to the size of the node in the Topology Visualization) and appends that to the Points.csv file to create the **Points_With_Degree.csv** file.

Visualization of the Internet Topology

Inputting the **Points_With_Degree.csv** and **Links.csv** files to Flourish website, we obtained a comprehensive visualization of the Internet Topology. We have submitted a image of the same as part of the deliverable that look as follows :



An interactive version of the Internet Topology Visualization that we obtained can be accessed using this link - Dynamic Visualization of the Internet Topology.

Findings about Internet Topology

We have made the following observations from the Internet Topology Visualization that we have obtained :

1. There are several instances where multiple nodes are connected to the same destination node, suggesting redundancy and robustness in the Internet network.

2. The nodes labeled as IXP (Internet Exchange Point) indicate the locations where multiple ISP's connect to exchange traffic. Nodes associated with global ISPs, such as Tata Communications, NTT America, and Telstra, play a crucial role in enabling global connectivity.

3. Some nodes have notably higher degrees, such as Google LLC routers and Facebook Inc. routers, which suggests that these entities serve as important hubs in the network due to their extensive connections.

4. There are nodes representing regional and local ISPs like Beam Telecom, Reliance Jio which are essential for providing connectivity at a more local level.

5. The number of hops that data packets traverse to reach the destination varies significantly for different combinations of sources and websites.

6. Some of the edges in the network have been repeated multiple times which indicates that good measures have been taken to enhance fault tolerance.

7. The presence of nodes from different countries and regions, including Singapore, Japan, Canada, and Sweden, highlights the global nature of the internet and its infrastructure.

8. The response times between hops are inconsistent and vary significantly from one hop to another as they depend on factors such as network congestion, routing efficiency, and the performance of the intermediate devices.

9. Some hops in the traceroutes correspond to major internet backbone providers like Hurricane Electric and Arelion Sweden AB. These providers play a crucial role in interconnecting different networks and regions.

10. We have also observed that Internet topology is not static, as traceroutes conducted at different times yield different results. This might be due to network maintenance, outages, and routing updates.