

Assignment 1

AI3020- Intro to Computer Networks
Indian Institute of Technology Hyderabad

Aakash Kamuju, Sumeeth Kumar, Lokesh Badisa

I. DELIVERABLES

A. *Visualization for Internet Topology*

Initially, we explored visualization tools like GoJS, NetworkX, Pyvis. We have finally settled with Pyvis because of its high interactivity. In Pyvis, the greater the number of connections between two nodes, the lesser is the distance between them. We have used the same symbol for routers, ISPs, IXPs, and different symbols for destinations and sources. We have given different edge colors for the nodes connecting to different destinations. We used GeoPandas for plotting the path visualization on a world map.

B. *Code and Script*

The raw data we collected was obtained by using traceroute and ipinfo, along with Python packages like ipwhois, ip2geotools, and socket. We used UNIX Regex (string splitting) to collect all the IP addresses from the traceroute output, and the terminal command ipinfo and Python package ipwhois were used to get information about each IP address. We used ip2geotools for the location of specific IP addresses, and the Python socket package was used to find the address of the source.

C. *Raw Data and Processed Data*

The raw data we generated includes IP addresses, AS Numbers, ranges of IP addresses, location of specific routers, and organization information. Private IP addresses didn't show the aforementioned properties as they are hidden due to being private. The private IP addresses are the routers within the same ISPs that don't respond to the traceroute command. They are also called Bogon IP Addresses.

We have 12 destinations from our 5 sources, which are 'ntt.com,' 'google.com,' 'archlinux.org,' 'Facebook.com,' 'lookup.icann.org,' 'Japantimes.com,' 'wikipedia.org,' 'News.com.au,' 'Cisco.com,' 'cloudflare.com,' 'Bbc.com,' and 'indiatoday.in.' The reason for selecting these websites is that some of them are native to specific countries but can be accessed throughout the world. We inferred that to reach these country-specific destinations, it needs to go through multiple servers, which need not be in the countries of source or destination. We have removed the processed private IP addresses in the data visualization. We also created a table with AS Numbers and their range of IP addresses and their respective organizations.

II. FINDINGS ABOUT INTERNET TOPOLOGY

A. *Content Delivery Networks (CDNs)*

- Many websites utilize Content Delivery Networks (CDNs) to efficiently serve their content. CDNs store website databases on strategically distributed servers worldwide.
- Prominent CDNs like Akamai Technologies Inc. and Fastly Inc. provide hosting services for various companies.
- CDNs contribute to reducing latency and enhancing the speed of content delivery.

B. Server Routing Variability

- There is no guarantee that the same server serves every user request. Routing depends on factors such as server/router traffic, request location, and network conditions. - Large companies like Google often bypass third-party services and host their own servers to maintain greater control over infrastructure and data.

C. Diverse Database Encounters

- During our traceroute analysis involving 5 different source points and 12 distinct websites, we encountered 26 different databases/servers. - This diversity highlights the intricate path that internet traffic can traverse before reaching its destination.

D. Intermediate Routing Behavior

- An interesting observation was the delay in serving requests by some servers. Instead of immediate service, servers occasionally redirect requests to other routers, relaying the request to the server. - This behavior could be part of a routing strategy to optimize traffic flow.

E. Unidentified Routers

- In our topology, we identified 73 routers whose roles as routers, Internet Exchange Points (IXPs), or Internet Service Providers (ISPs) remained uncertain. - Segregating these routers is possible by examining properties such as abrupt latency changes and the number of connected routers.

F. Wi-Fi Network Impact

- When connected to Wi-Fi networks, user requests pass through the local Wi-Fi network before entering the broader internet. This intermediate step in the routing process can influence the request's path.

G. Common Path Phenomenon

- Occasionally, packets from two different websites may follow the same path until a certain router, even beyond the local ISP. This phenomenon is called a 'Common Path' till a certain router.

H. AS Numbers and IP Addresses

- Multiple routers within the same ISP often share the same Autonomous System (AS) Number. - ISPs can operate with multiple AS Numbers, and each IP address within their range is associated with a unique AS Number.