

Visualizing the Internet Topology

Vishal Vijay Devadiga (CS21BTECH11061)

Satpute Aniket Tukaram (CS21BTECH11056)

Information

- What are AS numbers?

AS(Autonomous System) is a collection of IP prefixes under the control a single administrative entity. Each AS is assigned a ASN that is used in routing. Alloted ASN exceed 1,00,000 as of 2021.

- What are IP addresses?

IP(Internet Protocol) address is a label that is assigned to a machine connected to a network. It serves the purposes of: **Locating Addressing** and **Network Interface Identification**

- What do `traceroute` and `whois` commands do?

`traceroute` is a network diagnostic command that shows how packets routes between a source and its destination, also measuring the times in each hop. `whois` is a query that returns an internet resource's registered assignees.

Deliverables

Our submission has 6 folders:

- 1) Scripts: This contains all the scripts(python) used to extract and process data.
- 2) Trace Routes: This contains all the raw data of traceroute command.
- 3) Who Is: This contains all the raw data of the whois command.
- 4) Processed Data: This contains the processed data made by the scripts
- 5) Visualization: This contains the picture of the visualization of the internet topology
- 6) Report: This contains the report of the assignment.

Destination/Source Selection

We decided on a mix of Indian and American sites and a Russian site too. Following are the destinations:

- `www.iith.ac.in`
- `www.youtube.com`
- `www.ubuntu.com`
- `www.codeforces.com`
- `www.chess.com`
- `www.wikipedia.org`
- `www.github.com`
- `www.zomato.com`
- `www.olacabs.com`
- `www.zerodha.com`

Zerodha, Codeforces and Chess.com were hosted on cloudflare, sharing ASNs.

We have 5 sources:

- IIT Hyderabad Wifi
- Airtel Mobile Network
- Jio Mobile Network
- Looking Glass (New York)
- Looking Glass (Los Angeles)

Methodology

We executed the `traceroute` command with `-IAe` flags. We then processed this raw data to only include `AS routes`, total latency, IP Address, Destination and Source.

Using these IP Addresses of the processed data, we then executed the `whois` command, and found out the IP Range, and the Organization to whom the IP Address belongs to.

For the looking glass results, we manually scrapped all requirements from the raw data, as the output from the looking glass(*NTT*) was not consistant. Writing a script to process such data would have taken a lot of time having to remove all edge cases.

We then visualized all the available data in a diagram created on the `draw.io` site.

Results

Internet is **interconnected**. A lot of the routes have some common **IXPs**, facilitating the direct exchange of traffic between ASNs.

Even if two destinations seem unrelated, they could share a common routing path for a portion of their routes if it provides the most efficient and reliable transmission. For example, in our diagram, a lot of the routes went through either **RELIANCEJIO-IN** exchange point or **BHARTITELESONIC** exchange point before heading to the destination.

Also, a lot of sites are hosted on common hardware, to save costs. For example, **Zerodha**, **Codeforces** and **Chess.com** were hosted on cloudflare, sharing ASNs. Cloudflare was a performant host, having a lot less hops compared to other hosts.

Also, we observed that mobile networks often have their own routing hardware. For example, operators like **Jio** do not acknowledge **traceroute** and return only the starting point and the end point of the route.

Also the IITH router sometimes does not respond to **traceroute** command and leads to hopping between unknown endpoints.

Looking Glass does not give accurate latencies between the hops in the route. It was felt as if the actual latencies were magnitudes higher than the data being reported. Also looking glass could not find a route to IITH even after setting a higher maximum number of hops.