Depiction of Internet Topology

August 25, 2023

Authors:

- Abhay Shankar K: cs21btech11001
- Kartheek Sriram Tammana: cs21btech11028
- Kushagra Gupta: cs21btech11033

Files

This section contain the list of files included, as well as their purposes and their origin.

- `data_collection`: These scripts collect the data and generate the graph
  - `raw_traces/*.json`: Raw traceroute data (from `01-traceroute.py`)
  - `01-traceroute.py`: Run traceroute to all destinations (uses the `scapy` library)
  - `02-collect-ASNs.py`: Collect AS info for all IPs (uses a public API). Generates:
    * `ip_to_asn.json`: IP to ASN mapping (from `02-collect-ASNs.py`)
    * `as_info.json`: AS info (from `02-collect-ASNs.py`)
  - `03-clean-raw-traces.py`: Clean the traceroute data (collapse duplicates, add AS info, etc.)
    * Generates `clean_traces/*.json`: Cleaned traceroute data (from `03-clean-raw-traces.py`)
  - `04-build-graph.py`: Build the graph from the cleaned traceroute data.
    * Generates `graph.json`: Final graph (from `04-build-graph.py`)
- `viz`:
  - `src`:
  - `as_info.json`:

Execution

Building the graph

Data collection done by the `01-traceroute.py` script. It uses the `scapy` library to run traceroute to all destinations. It outputs a JSON file, and the collected JSON files from multiple different sources is available in the `raw_traces` directory.

To collect ASN info and build the graph, (using our given data):

```
cd data_collection
python3 02-collect-ASNs.py
python3 03-clean-raw-traces.py
python3 04-build-graph.py
```

Visualizing the graph

To run the visualization,

```
cd viz
npm install
npm run dev
```

then open http://localhost:5173 in the browser. Alternately, we have hosted it online at https://kst164.github.io/networks-assgn1/.

Presentation

Our final presentation is an interactive 3d graph, with nodes representing routers with the same AS number, and edges representing a link between two ASes.

We chose colleges across the globe as our destinations - some have hosted their websites locally, while others have used a third-party service or CDN. This gives us a clearer picture of the Internet.

Node properties:

- Each node represents one Autonomous System, and represent a collection of routers and switches.
- Upon hovering over a node, a popup displays the AS number, and the organisation that owns it.
- The nodes are colour coded based on the country of the organisation that owns the AS.
- Nodes are scaled in size based on their degree (i.e. number of incident edges). This gives a rough idea of the size of the AS.

Edge properties:

- An edge between two nodes represents a link between the two Autonomous Systems.
- Links between two routers within the same AS are not represented in the graph.
- Each edge contains packets - moving spheres travelling along the edge direction. The speed of a packet represents the latency of the link.
- The graph is a multigraph, i.e., multiple edges may exist between two nodes. This occurs when multiple traceroutes, from different sources/destinations, pass through the same link. This is not directly visible, and appears as packets of different colours, travelling at different speeds.

Global graph properties

- The whole graph is interactive, and a user can drag a node around and hover over a node to get additional data.
- Furthermore, the user may select a source or destination to highlight the nodes and edges constituting the path between them.
- The graph contains **48** nodes and **942** edges - some of these edges are between the same pair of nodes along different routes.

Takeaways

- The Internet is a vast and complex realm, with myriad interconnected networks and subnetworks and a grand superstructure. Through the controlled application of the `traceroute` tool, we have been able to map a portion of the Internet to a graph, through which we study its topology.
- Numerically, looking at the edge-to-node ratio, we can see that the graph is dense and well connected. Obviously, not all the edges of the graph correspond to different physical links, but the number is a viable yardstick for how well-connected different parts of the network are.
- The first tangible takeaway is the knowledge of Autonomous Systems and their layout throughout the internet. The various APIs and online databases relating AS numbers to IPs and organisations yields valuable knowledge about the contents of the Internet.
- It is immediately obvious from the graph that a large amount of traffic goes through several key AS'es, and that these AS'es are ISPs or Cloud platforms. Further investigation shows the likes of Bharti Airtel, Reliance Jio Infocomm Ltd., along with Fastly.net and Cloudflare, Inc. which are Cloud platforms.
- A more thorough review provides knowledge of ISP-client relations across the globe, including but not limited to the scope of the National Knowledge Network, exactly which ISPs service which organisations, and the fact that our college IITH is directly linked to several other IIT's directly without going through either NKN or Jio.
- The graph lends itself well to a deep investigation, which when conducted imparts several more insights. NKN itself appears to own two ASNs, viz. 55824 and 55847. However, when we check the ASN of IIT Guwahati, it shows 55847, indicating that this AS hosts some IIT websites. During our data colection, we found that IIT Kharagpur is also hosted in the same AS.