Visualization of Worldwide Internet Traffic and Routers

Thomas Kennedy
Southern Methodist University
tkennedy@smu.edu

Alec Siems
Southern Methodist University
asiems@smu.edu

Jacob Adkins
Southern Methodist University
jpadkins@smu.edu

Vipul Kohli
Southern Methodist University
vkohli@smu.edu

ABSTRACT

The motivation for this project is to provide an accessible representation of traffic throughout the Internet, which will clearly illustrate the distribution of routers and the usage/workload corresponding to each. This representation will be geographically aligned in a way so that the physical location of each router on the globe is correctly represented. Color coding and scaling thickness will aid in the aesthetic representation of this data.

1. RESEARCH METHODOLOGY

Our first step is to determine a set of source and destination IP addresses with which we can conduct our traceroute operations in order to gather relevant data for analysis. These addresses are retrieved using a script that parses and aggregates URLs from university websites worldwide. Our second step is to use a script of our own design to pull from the addresses and execute the traceroute command, returning the IP address of each router along the route from the source IP address to the destination IP address. Our third step is to use another custom script which utilizes a publically available API that can automate the process of determining the geographic location of all the collected IP addresses. Our fourth step is to put the data into an easily accessible format, as well as begin our interim draft.

Our fifth step is to determine a background graphic for the world map onto which we will project the router traffic infographic. During this step we will also develop a legend for the map, and finalize any stylistic choices. Our sixth step is to add the routers onto the map. Depending on their traffic and location, separate sizes and colors of icons will be used. Our seventh step is to add the connections between the routers to the map. Depending on their traffic and location, separate sizes and line thicknesses will be used. Our eighth step is to analyze and draw conclusions from the data we collected and the resulting map graphic. Our final step is to write the project report.

2. PREVIOUS WORK

William Cheswick and Hal Burch created the "Internet Mapping Project" in 1997 at Bell Labs, which since 1998 has collected traceroute data of more than a hundred thousand networks every day. While the project includes a visualization of the data, its main purpose is to display the growth of the internet over time. Our project will instead focus on the traffic and routers of the internet in its current state, and the projection of the data upon a map of the world, something the Internet Mapping Project does not include. [1]

Co-founder of Wired Magazine, Kevin Kelly, also started an "Internet Mapping Project", however his was a crowd-sourced image slide-show of how different people view the Internet. The images were a mix of conceptual, concrete, and comic. The project was a piece of artwork and not an informative map based on actual data. [2]

In February 2012, H.D. Moore attempted to send a ping to every IP address in the world. His results were surprisingly accurate and gave a good representation of available IP addresses across the globe. There were considerable downsides to his work, such as the complaints from law enforcement and hate mail from individuals not ecstatic about his work. Moore sent out over 3.7 billion outgoing messages, resulting in return messages from over 3 million destination IP addresses, which equates to over 2 Terabytes worth of data. Our timeframe and computing resources do not allow for this project to be as extensive as Moore's. [3]

Two possible resolution directions have appeared to improving the current Internet architecture's scalability problem: separation from the main transit core and elimination of the edge networks. We may see which is empirically implemented through our research. [4]

There is one popular tool that will help up locate the longitude and latitude of a particular IP address. The developer created a database of CSV files to help users pinpoint any IP address that is within the database. This will come in handy after we execute all traceroute commands and need to fill in our map. [5]

3. RESOURCES NEEDED

First, an appropriate list of URLs of the web sites of all university in the world is required from each continent. This list can be parsed from the page source of a web site we

---

[1] Cheswick, William. "The Internet Mapping Project"

[2] http://kk.org/mt-files/internet-mapping/index.html

[3] http://www.technologyreview.com/news/514066/what-happened-when-one-man-pinged-the-whole-internet/

[4] Meisel, D, et al. "Towards A New Internet Routing Architecture: Arguments for Separating Edges from Transit Core." http://conferences.sigcomm.org/hotnets/2008/papers/18.pdf

[5] http://dev.maxmind.com/geoip/legacy/geolite/

found (univ.cc). This HTML parser will most likely be self-programmed using JQuery. More research is required in either self-programming or finding a ip route data collection software that can iterate through several traceroute commands. After the traceroute ip routing data has been collected and dumped to a folder or text file, the major nodes will need to ranked, so a text file parser and analyzer is required which will need to be researched or self-programmed. After the nodes are ranked, an appropriate mapping software still needs to be found. Recommendations would be helpful for the mapping software. This software needs to support different colors of lines and markers as well as including a legend. ShareLatex will be the software used to write our interim and final report drafts.

4. RESEARCH PLAN

5 February 2015 - Research Proposal Due
12 February 2015 - Collected Destination URLs
19 February 2015 - Traceroute Source/Destination
26 February 2015 - Map Route IP Addresses to Location
5 March 2015 - Format Data, Interim Draft
5 March 2015 - Interim Draft Due
12 March 2015 - Choose Map Graphic and Develop Legend
19 March 2015 - Add Routers to Map
26 March 2015 - Add Connections to Map
2 April 2015 - Analyze and Draw Conclusions
9 April 2015 - Finish Project Report and Presentation