**An Experimental Visualization of the Internet Topology**

|  |  |
| --- | --- |
| 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 outcome of this experiment was a visualization of the topologies, media, and delays of communications across the major routing nodes of the Internet. This visualization defined peering between local Internet Service Providers (ISPs) and backbone ISPs as the infrastructure making global communication prolific. Since the nature of networking is economic rather than geographic, the resulting visualization takes an original design, highlighting the topology and connectivity of the Internet rather than just the location of its hosts. This paper experimentally analyzes how data gathered from a few Internet tools and registries was able to create a representative visualization of the global Internet structure. Among these tools are Traceroute, Whois, and BGPlay.

# INTRODUCTION

# RELATED WORK

Cheswick and Burch [1] collected Traceroute data of more than a hundred thousand networks every day since 1998. While their project included a visualization, their main purpose was to portray the growth of the Internet over time. Our experiment, instead, was to focus on mapping the Internet topology in its current state by specifically characterizing the major ISP nodes and their peering links, a concept beyond the scope of Cheswick and Burch’s research. However, we compared our findings to their work on a general level. #need to expand here after experiment#

In February 2012, H.D. Moore [2] attempted to contact every IP address on the Internet. Out of the 3.7 billion IP addresses contacted, over 3 million responses were received, amounting to over two terabytes worth of data. His results give a good representation of the number of IP addresses accessible to everyone in the world. However, our timeframe and computing resources did not allow for such an extensive data collection and analysis as Moore’s, so chose to limit our experimental scheme to the most meaningful IP addresses and URLs to draw the Internet to each country in the world.

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.[[1]](#footnote-1)

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.[[2]](#footnote-2)

3. EXPERIMENTAL METHODOLOGY

This sections describes the tools and procedure used to collect and analyze ISP and routing data in order to produce a visualization of the Internet to every country and continent which shows the topology, material, and routing involved.

#expand here

3.1 Data Collection and Analysis Software   
 In order to map the Internet to all of the countries in the world, a list of IP addresses of servers in each country in the world was required. Therefore, we chose a list of Uniform Resource Locators (URLs) of the Web sites of universities in every country provided by Univ.cc [3]. We targeted university Web sites over commercial Web sites since universities are more likely to have their Web servers in-house whereas commercial Web sites mostly outsource their server needs. Each country’s list from Univ.cc is proportional to the country’s area. For example, larger countries like Argentina has 83 URLs while a smaller country like Andorra has only 1 university.   
 After the URLs were collected, the next task was to determine the Internet routing to each of the university URLs. The IP2Location Traceroute [4] software converts these URLs to their respective IP addresses and determines the route to each of these IP addresses from the IP address of the chosen IP2Location host server. The IP2Location Traceroute application allows users to choose a server either in the United States (Phoenix, Dallas, or Los Angeles), Canada (Montreal), France (Paris), Germany (Dusseldorf), Netherlands (Amsterdam), England (London), Singapore, Japan (Tokyo), or Malaysia (Kuala Lumpur), a total of twelve servers in ten countries on three continents.   
 Traceroute.org [5] was another resource we explored to possibly run Traceroute originating in more countries, but preference was given to IP2Location to better suit the purpose of our experiment. Traceroute.org was disadvantageous since its Traceroute programs returned only each node’s IP address, URL, and hop delay (in milliseconds). However, IP2Location provided the same information as well as each node’s city, state, and country where the node’s host was registered. Noticing this, our aim became to originate the Traceroute from a few countries but reach all the countries with plans to trace the networks of the major local and global ISPs. By selecting origins in North America, Europe, and East Asia, we planned each destination university URL to be at most two hops away from at least one of the originating servers and much more distant from other originating servers to obtain both local and global paths.  
 After the Traceroute data was collected, the next task was to plot the nodes and links in a manner to determine the topology to effectively construct the final visualization. Graphviz [6] was best suited for this purpose.   
 Graphviz is a graphical visualizer of the DOT programming language engineered for several different operating systems including Windows, Solaris, Ubuntu, Fedora, and Mac OS (Leopard and later). Our DOT code used adjacency lists to represent unidirectional links from each originating node to each of its possible hops. Graphviz interpreted our DOT code to draw the corresponding directed graph, or web, with bubbles as the nodes and unidirectional arrows representing the links. Color coding is also included in both the DOT language and the Graphviz graphs to help create distinctions to extract Internet topology. Backbone nodes would have more connectors than the final destination nodes, so nodes having a defined minimum number of outgoing links or incoming links would be color-coded accordingly than the other nodes to make the topology possible.  
 The Center for Applied Internet Data Analysis (CAIDA) [7] rank several global ISPs along with their AS numbers. These AS numbers were applied in our experiment to Daniel Austin’s BGP Lookup Tool [8] to achieve an IP prefix which was then supplied to the BGPlay tool.  
 BGPlay is a JavaScript visualization tool that generates a graph representation of the links within five hops from the source IP prefix. Each of the linking nodes consist of the AS and ISP registration details which were investigated further.  
 With all the utilities of each software determined, the next task was to construct a bottom-up layered data collection and transformation scheme to process the different forms of routing and ISP data available to be able to concisely determine the topology,

3.2 Data Collection and Transformations

The first operation was to load all the university URLs from Univ.cc [3] to a text file to ignite our experiment. This was done by a Java program we composed to crawl through Univ.cc and save each country’s university URL list to the text file which would serve as the list of destination URLs for IP2Location.  
 The IP2Location Traceroute was a Web service which took in two HTTP POST parameters in its request header: the originating server location and the destination URL or IP address. After processing the request and completing the trace, the service returned a table in Hypertext Markup Language (HTML) with the route’s nodal data. Markup languages like HTML are very descriptive and hierarchical with their starting and ending markers. These characteristics of the language allowed for easy extraction and containment of data to the essentials which resulted in savings in terms time and storage.   
 We composed another Java program to iterate through the university URL list and execute the corresponding Traceroute requests originating from all twelve server locations which resulted in about 25 thousand routes being collected with reachability to all of the seven continents.

1. 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 [↑](#footnote-ref-1)
2. http://dev.maxmind.com/geoip/legacy/geolite/ [↑](#footnote-ref-2)