Demo: Interactive Access to Internet Topology Data

Kc Claffy, Bradley Huffaker, Young Hyun, and Marina Fomenkov*;
Center for Applied Internet Data Analysis, University of California San Diego
*CAIDA, SDSC, UCSD, La Jolla, CA, 92093-0505, USA; email: marina@caida.org

Abstract: *CAIDA researchers have gathered the largest set of network topology data used for a broad spectrum of scientific research: from physics to biology, from cyber infrastructure vulnerability assessments to theory of complex networks. In order to broaden the research community access to the wealth of raw and curated data that we produce, we are developing two interactive portals, AS Rank and Vela, that we would like to demonstrate at the Gateways Conference.*

1. Introduction

The Center for Applied Internet Data Analysis (CAIDA) [1] based at the San Diego Supercomputer Center (SDSC) at the University of California San Diego has been trailblazing the field of measurement and analysis of the Internet topology since 1998, and the Internet research community has grown to rely on CAIDA as one of the best sources of available data about the Internet infrastructure. We have deployed and maintain a globally distributed measurement platform Archipelago (Ark), which consists of more than 200 monitors (and growing) deployed on 6 continents and controlled by a central server at CAIDA [2]. We designed and implemented the Ark platform, optimizing its functionality to support scientific experiments that require coordinated active measurements of the global Internet. One measurement experiment continuously running on Ark is macroscopic Internet topology mapping: CAIDA tracks global IP level connectivity by sending probe packets from Ark monitors to millions of geographically distributed destinations across the Internet address space. Over the years, we have accumulated over 43 billion traces (path measurements), stored in 20 TB of Internet topology data. These data sets deepen our insight into the structure, behavior, and evolution of the global Internet, and also enable modeling and simulation of malware propagation and containment measures, infrastructure stability and vulnerability assessments, longitudinal studies of Internet topology evolution, and Internet address mapping and inferences.

Presented at Gateways 2016, San Diego Supercomputer Center, La Jolla, CA, November 2-3, 2016.
https://gateways2016.figshare.com/

2. Interactive Access to CAIDA Internet Topology Data

CAIDA researchers created and support two online services for users to interact with our data.

1. AS Rank

The Internet is composed of tens of thousands of Internet Service Providers (ISPs) that build and operate individual networks known as Autonomous Systems (AS). ISPs voluntarily engage in cooperative relationships to collectively and ubiquitously route traffic across the global Internet. These relationships manifest as business agreements that translate into engineering constraints on traffic flows within and across networks in the Internet routing system. We developed a methodology to construct an AS-level graph of the Internet from publicly available routing data, annotate nodes in the graph according to economic attributes of networks, e.g., content, transit, and annotate links with CAIDA-inferred economic relationships between networks, e.g., customer, provider.

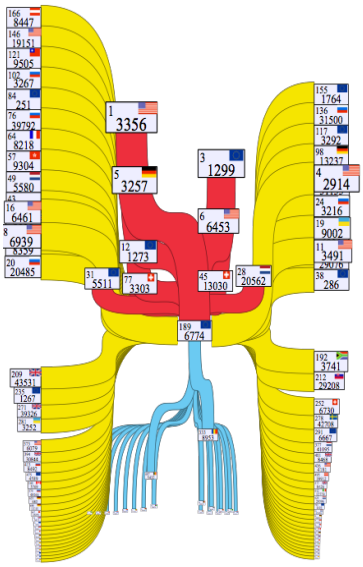

Figure 1 Connectivity relationships for one AS.

The inferred AS relationships form the basis of the AS Rank [3] service that delivers a rank-ordered list of Autonomous Systems (AS) derived from Internet routing data (Border Gateway Protocol, or BGP data), registry data (WHOIS), address ownership, and transit metrics associated

with each AS. The interactive AS Rank portal allows the user to view the analyzed and annotated data at various granularities, and to drill down on a single AS to see its customers, peers, siblings and providers (Figure 1) as well as geographical information about its neighbors (Figure 2). AS Rank also enables vetted operators to enter corrections to identify and correct false inferences, and to increase the precision and coverage of geographical annotations. This service represents the state-of-the-art with regard to the completeness, accuracy, and richness of the macroscopic Internet maps we provide to the research community.

2. *Henya: Topology Data Querying*

The Ark platform has operated since 2007 (succeeding CAIDA's previous *skitter* infrastructure). The collection of topology data (20 TB currently) continues to grow by approximately 316 GB (766 million traces) per month, or 9 billion traces per year. To support efficient retrieval and analysis of data of interest from this enormous collection, we are developing a topology query system – Henya. The goal of this querying system is to allow users to explore topological properties of a massive archive of path measurements. Example queries might include requesting all paths that pass through (or alternatively, reach) a specified set of IP addresses, prefixes, ASes, or countries. In addition to command-line access to this vast data archive, which allows researchers to execute queries and write their own analysis and visualization scripts, we also support simplified interactive access via a web interface (Vela).

The Vela portal, currently open to academic researchers, combines two main functions. It enables ad hoc interactive topology-on-demand experiments on the Ark platform: users can conduct customized Internet probing and path measurements from any Ark monitor, using various Internet communication protocols as needed to explore different aspects of Internet behavior.

Second, Vela supports functionality to query, process, analyze, and visualize historical traceroute (i.e., path measurement) data. The goal of these interactive analysis and visualization capabilities is to increase accessibility of the measurement data and offer novel insights to a variety of disciplines and domains. In the next stage of development, we will implement commonly desired queries, analysis, and visualization options.

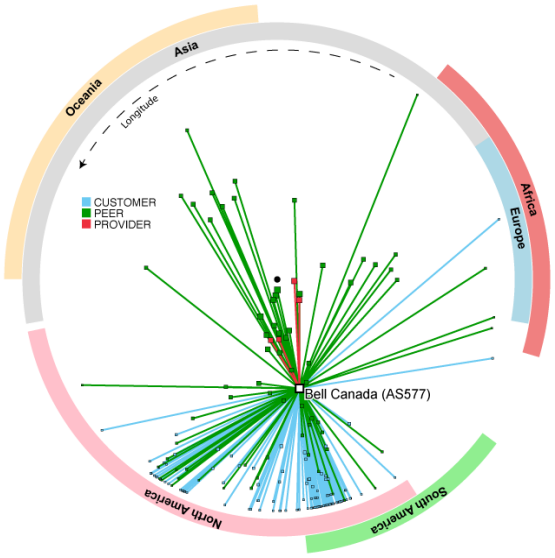

Figure 2 geographic footprint of a single AS.

3. Conclusion

We offer two interactive portals enabling access to Internet topology data collected by CAIDA over a nearly 20-year period. This is work in progress. We continue to add Ark nodes to fill gaps in our global measurement footprint. We also continue to extend and improve the AS Rank service, and we plan to broaden access to Henya catering to researchers, operators and the long tail of casual users. We look forward to further improving the data's usability, as well as the functionality and convenience of the portals to access it, with the help of the Science Gateways Community Institute.

4. Acknowledgments

This work is supported by National Science Foundation (NSF) grant CNS-1513283. This material is also based on research sponsored by DHS S&T/ HSARPA/CSD, BAA 11-01-RIKA and Air Force Research FA8750-12-2-0326. The views and conclusions contained herein are those of the authors and do not represent official policies or endorsements, of NSF, DHS, AFRL or the U.S. Government.

5. References

1. http://www.caida.org/
2. http://www.caida.org/projects/ark/
3. as-rank.caida.org
4. http://www.caida.org/projects/ark/vela/