Generating Realistic Internet Topologies Using Internet Topology Measurements

Peter Boothe

with thanks due to
Andrzej Proskurowski, Art Farley, Dan Stutzbach, Jun Li, and more

What is the Internet graph? This graph represents the inter-network connections that make up the Internet. Every vertex is an independent network (autonomous system or AS) and every edge represents a contract between two ASs. There are two forms of contract - peering, which is a bidirectional edge, and customer-provider, which is an edge from the provider to the customer. In the valley-free model of Internet routing, packets can use edges regardless of their direction, but no packet can travel against an edge after it has traveled with an edge.

[1] T. Erlebach, A. Hall, A. Panconesi, and D. Vukadinovic. Cuts and disjoint paths in the valley-free path model. In *Combinatorial and Algorithmic Aspects of Networking*, 2004.

[2] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, 2001.

[3] Z. Beerliova, F. Eberhard, T. Erlebach, A. Hall, M. Hoffmann, M. Mihalak, and L. S. Ram. Network discovery and verification. In *Proceedings of the 31st International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2005)*, volume LNCS 3787, pages 127–138. Springer, 2005.

[4] Peter Boothe, Zdeněk Dvořák, Art Farley, and Andrzej Proskurowski. Graph Covering via Shortest Paths, *Congressus Numerantium* (to appear)

Problem

Even though it's entirely human-made, we don't know the topology of the Internet. Knowledge of the Internet topology would be of great use to network researchers, ISPs, and policy makers. Can we recover the Internet topology from available data?

Available Data

Routeviews and RIPE have archives of routing tables for a select number of monitors placed throughout the Internet. Each routing table provides a list of shortest valley-free paths from the monitor to every other network in the Internet, and Gao's algorithm assigns a direction to each edge. [1,2]

Proposal

Generate a set of graphs, each of which, if monitored in exactly the way the Internet was monitored, will yield the measured shortest valley-free paths. By sampling from this set, we can provide a variety of plausible Internet topologies.

Problem

How can we take a list of shortest paths* and characterize the set of graphs that these shortest paths have come from?

* this argument can be extended from shortest paths to valley-free shortest paths, and we use this extension in practice

Graph Theory

If we have a shortest path of a graph, then ALL edges on that path must be in the graph, and NO edges in the graph connect any vertices that are more than two hops apart. Thus, every shortest path provides us with an assertion of edges and an assertion of non-edges. [3,4]

Set of Graphs

Our set of possible graphs is the set of all connected subgraphs of the complete graph that contain all 63,001* measured edges and all 26,521 vertices and none of the 1,241,476 forbidden edges. There are $10^{211,332,823}$ such graphs.

* Numbers from Routeviews data for 22 Sep 2007

Problem

This is a very large set of graphs, and some of these graphs seem, a priori, to be very unlikely. For example, the Internet graph is thought to be sparse, but many of the possible graphs are not, How can we narrow this set down?

Statistics

If we assume that our set of monitors has average degree equal to the average degree of the Internet graph, then the number of graphs shrinks to $\sim 10^{23,664,454}$. Stronger assumptions about the representativeness of the set of monitors can lead to further reductions in size.

Solution

Sample the set of possible graphs which have the same average degree as the set of measurement points, or, more restrictively, the same degree distribution as the set of sample points. If a given result is expected to hold over this sample set, then it is likely to be true for the Internet graph.