Analyzing the AS Graph Instead of Just AS Graph Measurements

Peter Boothe

31 October 2007

Abstract

Measurement of the AS graph is usually performed on a subset on the subset of the AS graph that is captured with measurements. In this paper we describe a method using AS graph measurements to derive a family of possible AS graphs, each with a described likelihood.[1]

1 Introduction

Even though it is entirely human made, we do not know the topology of the Internet. Knowledge of this topology would be useful for protocol designers, policymakers, and Internet Service Providers. Because the Internet is an IP network, and IP networks are built in layers, we must be very careful about exactly which layer we are talking about at any given time. At one level, the Internet is a connection of independent networks (autonomous systems or ASs) who have chosen to interconnect with each other and carry each other's traffic to provide mutual connectivity. It is these agreements that make the Internet an inter-network, and so we focus on this layer as our object of study.

The AS graph is the layer of Internet topology in which each independent network is modeled as a single vertex, and a link from one vertex to another means that there exists some contract between the two networks to directly exchange traffic. These contracts take two forms: peering and customer-provider. Peering links are where two autonomous systems agree to exchange traffic as equals, while customer-provider links imply one AS paying another AS money to carry traffic and provide connectivity. In the AS graph, we model links as directional from customer to provider — in effect, the arrows "follow the money". Peering links are therefore bidirectional, as no money changes hands. In the **valley-free model** of Internet routing, introduced by Gao in 2001[2], a packet's route consists of going in the same direction as some number of edges (possibly zero), and then going against some number of edges (again possibly zero). With this model of routing, we capture the idea that not all edges are useful for all autonomous systems.

Traditional analyses of the AS graph have involved downloading as many routing tables as possible (usually from RIPE[3] and Route Views[4], possibly from other sources as well) and then collating all of them into a single unified view that is assumed to be as complete as possible. This is in direct conflict with the fact that it is known that a significant number of edges are missing from these views. In this paper we outline a method of deriving possible and probable AS graphs from the measured data. Using this technique, it should be possible to derive, with measured certainty, measurements of the actual AS graph instead of simply taking the bounded AS graph and assuming it is complete. This is very important, because even simple graph measurement such as degree distribution have been show to be non-representative when a graph is sampled by shortest paths [1].

2 Proposed Pipeline

For a desired network statistic (any function whose input is a graph and whose output is a single number), we propose to measure that statistic in the following way:

1. Create a combined list of shortest valley-free paths from all available sources of routing data
2. Assign a direction to every edge using Gao's algorithm.[2]
3. Enumerate all edges which can not exist, because their existence would contradict our measurements being shortest valley-free paths.
4. Repeatedly, until the desired level of precision is reached, do the following:
    (a) Sample from the set of graphs that contain all measured edges and no impossible edges

---

[1] 2-page abstract submission for NIPS'07 workshop on statistical models of networks. `http://www.cs.ubc.ca/~murphyk/nips07NetworkWorkshop/`

   (b) Refine the sample until it is a sample with high likelihood
   (c) Measure the desired network statistic on this sample graph
   (d) Record both the value of the network statistic and the likelihood of the graph that the network statistic came from

5. Coalesce all measurements and likelihoods into an expected value and standard deviation.

Once this process is complete, we should have a good idea of what the value for a given network statistic on the AS graph actually is instead of merely what the network statistic of the measured subset of the AS graph is. In this description of the pipeline, the steps with the most hand-waving are steps 3 and 4b, so we devote the rest of this abstract to discussing them in more detail.

3 Impossible Edges

On a valley-free shortest path, for every pair of vertices A and B on this path that are separated by at least one vertex, one of the following situations must apply: Either $A \rightarrow \ldots \rightarrow B$, from which we conclude that there can be no edge from A to B; $A \rightarrow \ldots \leftarrow B$, which implies that there is no edge from A to B or from B to A; or, finally, $A \leftarrow \ldots \leftarrow B$, which implies there is no edge from B to A. All of these conclusions follow directly from the fact that the given path is a shortest valley-free path, because if any of those edges existed, there would exist a shorter valley-free path.

For a path of length k, this insight allows us to mark $O(k^2)$ edges as impossible. If we perform this algorithm on every single measured path, we can enumerate all impossible edges. Thus, we randomly choose a graph that is a supergraph of the union of all measured edges, but is a subgraph of the complete graph with all impossible edges removed. From this, we can generate a possible graph. Unfortunately, for graphs the size of the AS graph, there is an astronomical number of possible graphs, and not all graphs are equally likely to be the measured AS graph. This brings us to our next technique, which uses ideas from phylogenetic tree reconstruction in an effort to refine our randomly chosen possible AS graph into a more probable AS graph.

4 Refining the Sampled Graph

If we assume that our collection of measurement points is representative, then we can assign a likelihood to a possible AS graph by evaluating how likely it is that the degree distribution of our measurement points is a sample taken from the degree distribution of the possible AS graph. We can further refine this likelihood by comparing not just degree distribution, but also average shortest valley-free path length from our measurement points and other parameters. The main insight is that, for our measurement points, there are some aspects which we know completely. One of these aspects is the degree of each measurement point, and another is the valley-free distance from the measurement point to every other vertex in the graph.

Using those aspects which we know with certainty, we can create a distribution of their values over our sample set and a distribution of these values over the candidate AS graph, and then use standard statistical techniques to evaluate how likely it is that the sample distribution came from the proposed AS graph's distribution. Armed with this technique, we can then use hill-climbing methods to alter our possible AS graph into a more probable AS graph.

5 Summary

In this abstract we outline a proposed algorithm for measuring the properties of the AS graph and for gauging the uncertainty of our measurements. The next step, which is in process, is to implement these methods and to measure various aspects of the AS graph. It is also hoped that these insights into dealing with incompleteness in graph data can be useful in other domains where graphs are sampled and only a subgraph is known and we want to measure global properties of the graph.

References

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling; or, power-law degree distributions in regular graphs. In *Symposium on Theory of Computing*, May 2005.

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

[3] The RIPE Routing Information Services. http://www.ris.ripe.net, 1999-.

[4] University of Oregon Route Views Project. http://routeviews.org, 1997-present.