Understanding the Effect of P2P Overlay on the AS-level Underlay

Amir H. Rasti
amir@cs.uoregon.edu
University of Oregon

Reza Rejaie
reza@cs.uoregon.edu
University of Oregon

Walter Willinger
walter@research.att.com
AT&T Labs-Research

1. INTRODUCTION

During the past few years, Peer-to-Peer (P2P) applications have become increasingly popular over the Internet. In these applications, a large number (*e.g.*, millions) of geographically distributed end-systems (or peers) form an overlay to exchange content and share their resources. Several recent studies have observed the network traffic at one or few vantage points and shown that P2P applications significantly contribute to the overall Internet traffic(*e.g.*, [2]). Given the distributed nature of P2P applications, the observed traffic associated with a P2P application across different regions (*i.e.*, ASes) in the Internet depends on several factors including *(i)* the location of participating peers, *(ii)* the connectivity among the peers, and *(iii)* the BGP routing of traffic among ASes in the underlay. Therefore, the observed traffic for a P2P application at a single (or few) vantage point(s) clearly does not represent the extent of its impact across the network. Capturing an accurate view of the global impact of a P2P application requires access to a large number of vantage points that is prohibitively expensive. To our knowledge, none of the previous empirical studies on P2P systems have characterized the global impact of a P2P application on the AS-level underlay.

This extended abstract describes our ongoing project on assessing the impact of a given P2P overlay on individual ASes in the network (*e.g.*, offered load). Toward this end, first we present a simple and effective methodology to translate a P2P overlay into its corresponding AS-level underlay. In such an underlay, peers within individual ASes are grouped and each connection of the overlay is represented by its corresponding AS-level path. Second, we briefly present the array of analysis that can be conducted using the derived AS-level underlay, in order to characterize the impact of the overlay on the network from different angles. Finally, we present some of our preliminary results.

*CoNEXT'*07, December 10-13, 2007, New York, NY, U.S.A.

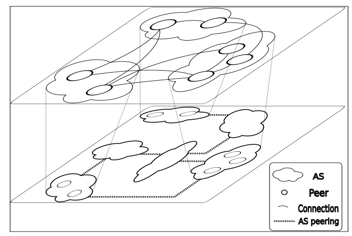

Figure 1: The overlay and the underlay

2. METHODOLOGY

Our methodology to assess the impact of a P2P overlay on the network includes the following high level steps (see Figure 1): *1) Capturing a P2P Overlay*: We capture a snapshot of a widely deployed P2P overlay (*e.g.*, Gnutella) which contains the participating peers and all the pairwise connections among them.

2) Grouping Peers within Individual ASes: We use snapshots of BGP routing tables from a close-by date (that are archived by the RouteViews project) to identify the corresponding AS for each peer in the overlay, and then group all the peers within each AS. The AS-grouped overlay clearly shows what fraction of overlay connections connect peers within individual ASes or peers in different ASes.

3) Inferring the AS-level Underlay: We derive the AS-level path for all connections of the overlay that connect peers in different ASes. Empirically determining the AS-level path between an arbitrary pair of ASes is complicated and prohibitively expensive. To cope with these challenges, we use the C-BGP simulator [3] to determine the AS-level paths. This simulator uses an inferred AS map with AS peering relationship, translate the provided AS relationships into common routing policies, and determines valley-free AS-level routes between two given ASes. We use the inferred and annotated AS-level topologies that are provided by CAIDA [1].

It is worth noting that inferring AS-level connectivity and determining the relationship between ASes remains a difficult problems and are still active areas of research. All the inferred and annotated AS maps, including those used here, come with a number of caveats and have to be taken with a grain of salt. While our results may not accurately quantify the effect of a P2P overlay on the network, but we believe that the presented trends are valid and provide a very valuable insight on the overall impact of the P2P application on the network.

Estimating Traffic: The derived AS-level underlay clearly shows the ASes that are crossed by (*i.e.*, carry traffic for) individual connections of the overlay. Since we can not capture actual traffic on individual connections, we use the *betweenness* of individual connections in the p2p overlay as a relative weight to assess the volume of traffic they carry compared to other connections of the overlay. Betweenness of a link represents the number of shortest paths in a graph that the link is located on.

3. ANALYSIS

The derived AS-level underlay allows us to conduct an array of analysis to characterize the impact of a P2P overlay on those ASes that host the corresponding peers (*i.e.*, edge ASes) and those ASes that only carry traffic (*i.e.*, core ASes). We leverage the ranking of each AS that is provided by CAIDA as well.

View from the Edge ASes We capture the impact of a given overlay on the edge ASes using the following metrics: *(i)* the distribution of peers across edge ASes, *(ii)* correlation between the number of peers in each AS and its CAIDA ranking, *(iii)* the percentage of edges that are located within a single AS, and *(iv)* distribution of AS-level path length across different connections, *(v)* the identity of top-N ASes that host the largest number of peers.

View from the Core ASes: We use the following metrics to characterize the impact of a P2P overlay on core ASes: *(i)* the distribution of transit load across individual core ASes, *(ii)* the correlation between their load and the CAIDA ranking, *(iii)* the identity of top-N ASes that carry most of the traffic.

We also use some of the existing techniques to estimate the tier of each AS. Using this rough information, we can estimate how far the generated traffic by a P2P application propagates through the AS hierarchy. We are repeating this analysis using our data-set over the past 4 years to investigate changes on the impact of these P2P applications over this period.

The next phase of our analysis is to examine the effect of peer connectivity and peer location on the resulting impact on the AS level underlay. Toward this end, we will change the identity of peers as well as their distribution across different regions (N. America, S. America, Asia, Europe and Africa). We plan to use a large pool of IP addresses for peers that we have captured as part of various P2P applications such as Gnutella, BitTorrent and Kad.

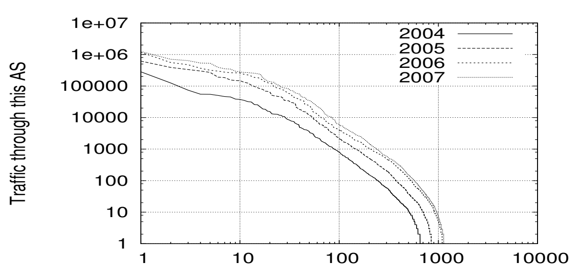

Figure 2: Gnutella connections passing through as AS

Top 10 Core ASes in '04		**Top 10 Core ASes in '07**	
ASNO	Name	ASNO	Name
1668	AOL	7018	AT&T
7018	AT&T	1668	AOL
174	Cogent Comm.	3257	Tiscali(DE)
2914	Verio	13237	LambdaNet(EU)
7911	Williams Comm.	174	Cogent Comm.
2828	XO Comm	3356	Level 3
3356	Level 3	812	Rogers Cable
209	Qwest	3602	Sprint(CA)
3257	Tiscali(DE)	209	Qwest
19548	Adelphia	19548	Adelphia

Table 1: Top 10 Core ASes by Gnutella traffic

4. PRELIMINARY RESULTS

Figure 2 shows the distribution of the number of Gnutella connections that crosses individual ASes for four different Gnutella snapshots that have been captured once every year since 2004. During this three year period, the size of the Gnutella network has changed in many ways (*e.g.*, the two tier architecture has been adopted) and tripled in size. Figure 2 shows a couple of interesting points: First, despite the dramatic increase in network size, the number of ASes that carry a significant fraction of traffic (*i.e.*, more than 100 Gnutella connections) is relatively small, *i.e.*, less than 500 ASes. Second, the distribution of load across affected ASes has not significantly changed over time and appears to be heavily skewed where a small number of ASes carry a large fraction of traffic.

One interesting question is whether the identity of ASes that carry most of the traffic has changed over time? Table 1 shows the identity of top-10 ASes that carry most of the Gnutella connections in 2004 and 2007 snapshots. This table shows that 7 of the top-10 core ASes have not changed during the last three years despite the dramatic increase in network size. However, the contribution of different core ASes and thus their ranking has changed.

We are currently conducting a wide range of analysis on a set of snapshots that span over four years in order to provide deeper insight on the impact of P2P applications on the AS level underlay.

5. REFERENCES

[1] Cooperative Association for Internet Data Analysis (CAIDA). http://www.caida.org/, 2007.
[2] A. Parker. P2P in 2005. http://cachelogic.com/research/2005_slide01.php, 2005.
[3] B. Quoitin. C-BGP. http://cbgp.info.ucl.ac.be/, 2007.