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

Amir H. Rasti, Reza Rejaie
University of Oregon

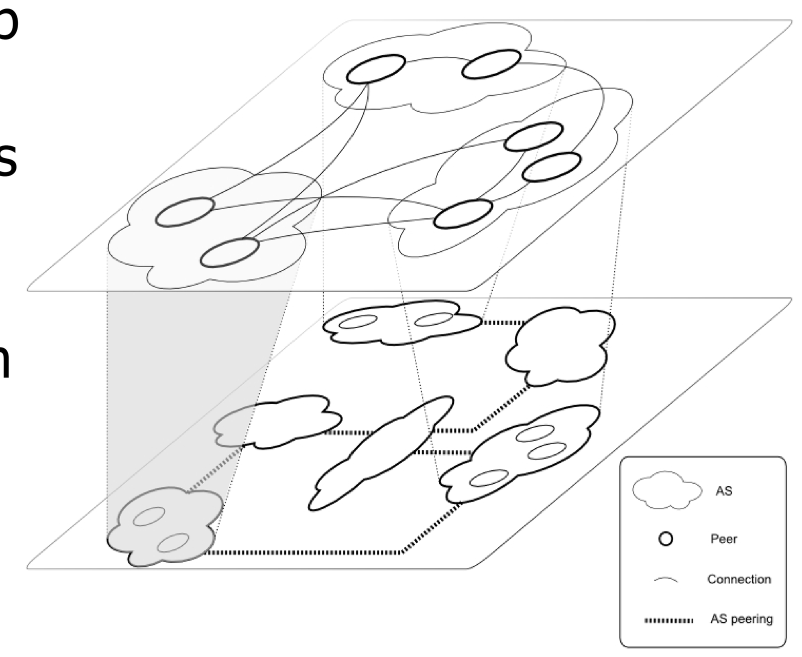
Walter Willinger
AT&T Labs - Research

1. Introduction

- **Why is it important?**
  - P2P systems have become very popular in a wide range of applications.
  - Each P2P application usually hosts a large number of users with a wide geographical distribution.
  - Some measurement studies performed at limited vantage points show that a large portion of the Internet traffic is associated to p2p applications.
- **Target questions:**
  1. For a given overlay, how can one capture and represent its effect on the underlying network?
  2. How do changes in the overlay affect the underlay ?
- **Factors** potentially affecting the amount of P2P traffic on each AS:
  1. Overlay topology and peers' connectivity and location
  2. Pattern of traffic generation and routing in the overlay.
  3. Underlay topology and routing.

2. Methodology

1. Capture **topology snapshots** of some popular p2p application:
   - Gnutella top layer overlay taken during the years of 2004-2007 using *Cruiser*. (177k–1.2M peers)
2. In each snapshot, **group** the peers based on their originator AS (according to BGP data gathered from *RouteViews*) (1872-3726 ASes)
3. Obtain **AS relationship** snapshots from *CAIDA*, one for each p2p snapshot taken close (within a month of) each.(38k – 50k relations)
4. Run ***C-BGP*** over each AS relationship snapshot to determine the "valley-free" AS-path carrying traffic between each pair of connected peers (100k-400k)
5. Calculate the traffic load on each AS for the following assumptions:
   a. Equal bidirectional traffic on all p2p connections
   b. Equal traffic generation at all nodes
      - Estimate the traffic of each connection using *betweenness* of each edge in the overlay graph.



Underlay Simplification

- AS represented as single node
- Incomplete AS relationship information
- Simplified AS relationships
  - Customer-Provider
  - Peer-Peer
  - Sibling-Sibling

Analyses Performed

1. Base case
   - 4 Gnutella snapshots
2. Randomized Connections
   - Randomize all connections keeping the degree of each peer fixed.
   - Any change in the underlay load pattern would show the effect of peer connectivity pattern
3. Random Peers
   - Using the pool of peers from a Gnutella snapshot, generate several random overlays with the same degree distribution as the original
   - Effect of peer identity and location
4. Regional Proportions
   - Using the same pool, generate random overlays with different regional proportions.
   - Effect of peer location

Tab 1. Top Core ASes Ranking

Name	Tier	07	Random Conn(Δ)	Random Peers(Δ)
AT&T(US)	1	1	0	0
AOL(US)	2	2	0	0
Level3(US)	1	3	0	0
Tiscali(IT)	2	4	0	0
LambdaNet(DE)	2	5	0	0
Cogent(US)	1	6	0	0
TeliaNet(SE)	2	7	-1	0
Sprint(CA)	2	8	+1	0
Rogers(CA)	2	9	0	-1
FranceTelecom(FR)	2	10	0	+1
EduNet(US)	2	11	0	0
Server Central(US)	2	12	-1	0
Sprint(US)	2	13	-1	0
COLT(NL)	2	14	-1	0
QWEST(US)	1	15	+3	0

3.Metrics and Results

Metrics

- Path length distribution
- Distribution of #connections carried per AS
- Distribution of top tier:
  - Diffusion in hierarchy
- Identity of top core ASes

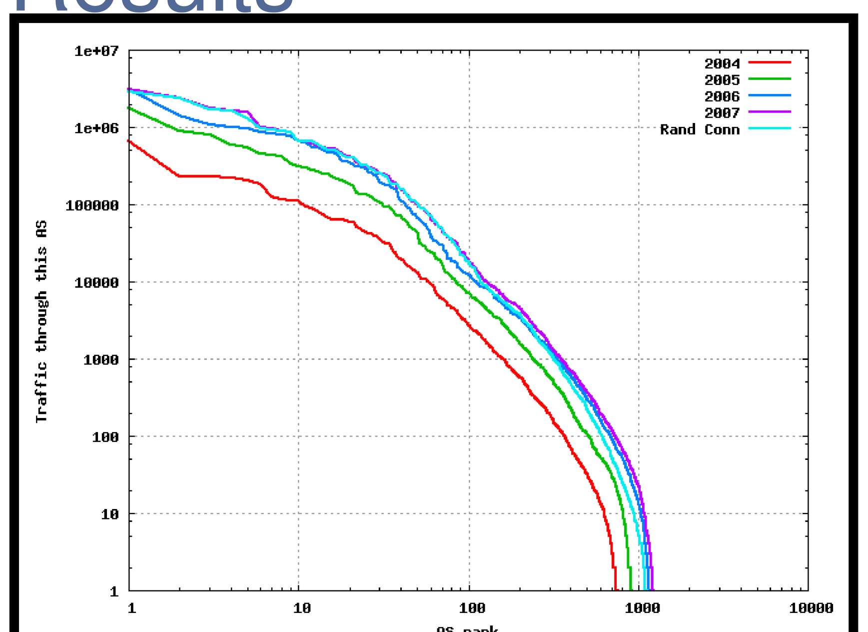


Fig 2: Distribution Of Transit Traffic

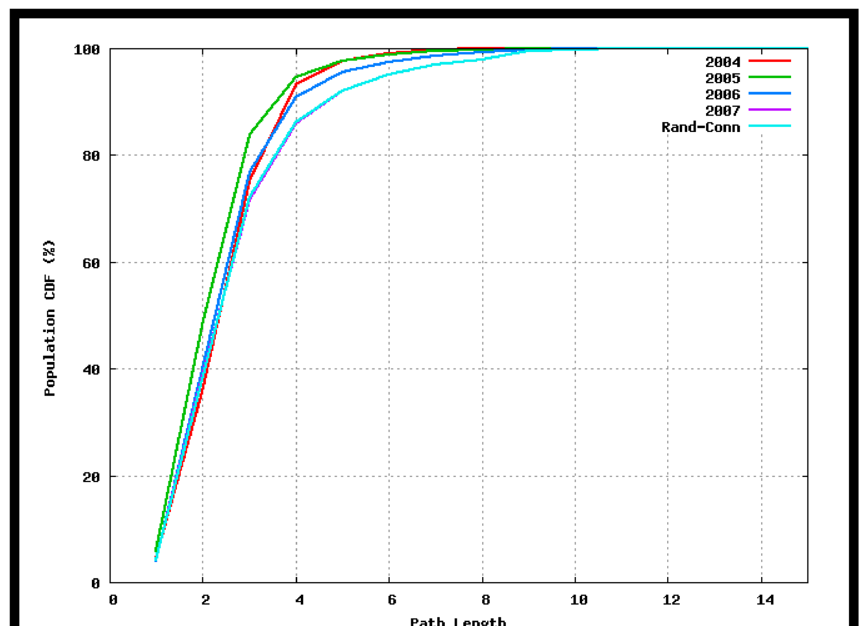


Fig 3: AS Path length distribution

%	04	05	06	07	R.C.
T1	84	73	64	63	63
T2	16	27	36	37	37
T3	0	0	0	0	0

Tab 2: Distribution of path top tier

4.Ongoing Work

Any-Any analysis

- Gather a large pool of peers from several recent snapshots (Gnutella & Kad)
- Divide ASes to 3 main grps of NA, EU, AS
- Assume full-mesh connectivity
- Separately, perform the analysis for all P2P connection from zone X to zone Y

Sub-path analysis

- Calculate the load on all sub-paths
- Determine the value of each AS

Traffic Matrix Analysis & Modeling

CoNext, New York, December 2007