Deploying Efficient Internet Topology Primitives

Naval Postgraduate School
Robert Beverly

September 18, 2013
Team Profile

• Naval Postgraduate School:
  – US Navy’s Research University
  – Located in Monterey, CA
  – 1500 students (all 5 services, civilians, foreign military)

• Our team:
  – **PI**: Robert Beverly
  – **Faculty**: Geoffrey Xie (NPS CS), Ralucca Gera (NPS Math), Arthur Berger (Akamai)
  – **Students**: Guillermo Baltra, Billy Brinkmeyer, Daryl Lee, Sam Trassare
Customer Need

- **Internet-scale Topology Mapping**
- **Need:**
  - Topology of Internet remains poorly understood
  - Critical infrastructure protection: robustness, vulnerability, correlated failures, IPv4/IPv6 interdependence, etc.
  - DHS BAA: “...identify infrastructure components in greatest need of protection.”
  - Researchers: modeling, prototyping new protocols, clean-slate designs, Internet evolution, etc.
- **Production systems, e.g. Ark, iPlane:**
  - Require $O(\text{weeks})$ to map
  - Induce significant load
  - Can miss short-lived events (which may be of most interest)
Approach Summary

• Started with theory primitives we proposed in [BBX10]
• Key Insights:
  – Utilize available external knowledge
  – Maintain state over prior rounds of probing
  – Adaptively sample to discover subnet structure
  – Maximize probing efficiency and information gain:
    • Which destinations to probe
    • How/where to perform the probe
• Implement in production on CAIDA’s Archipelago (Ark)
• Gather performance metrics
A Performance Metric

- Hard: how to evaluate “quality” of inferred topologies?
- Developed *edge/vertex symmetric difference (esd/vsd)* metric:
  - Intuitive (0-100%) difference between two topologies
  - Fast, scalable

\[ \text{vsd}(G, H) = \frac{|V(G) \setminus V(H)| + |V(H) \setminus V(G)|}{|V(G)| + |V(H)|} = \frac{1 + 1}{6 + 6} = 16.7\% \]
Utilize External Knowledge

- System input is set of global BGP prefixes (e.g. routeviews)
- Use knowledge of how networks are commonly provisioned and subnetted:

Easier to believe A and B in different subnets

Even though same “distance” apart. Use Least Common Prefix [BBX10]: find destinations with best probability of being in different subnets
Adaptive Probing

- Binary search each prefix, prune leaves that do not returning new topology
- Maintain set of interfaces discovered within the AS advertising the target prefix
- If new interfaces discovered by a probe, subdivide prefix and probe sub-prefix

- Design based on real-world challenges implementing primitive in [BBX10]:
  - No edit distance (distorted by load balancing)
  - Not pair-wise, no longer memoryless
  - Permits different vantage point for each probe, thus enabling integration with vantage point spreading
Maintain State

- We find 50% of prefixes probed by only ~10 monitors
- Thus, the choice of vantage point matters
- Developed and implemented *Ingress Point Spreading*:
  - Examine the set of ingresses into the target network discovered during prior rounds of probing
  - Rank order vantage points per target network to exploit ingress diversity
  - Expansion to “notional ingresses” permits any number of vantage points to be rank ordered intelligently
  - Prevents premature termination of adaptive sampling algorithm
Benefits

- Probing 50,000 randomly chosen BGP prefixes
- Compared to state-of-the-art Ark system
- More topology with half the load and time

<table>
<thead>
<tr>
<th>Metric</th>
<th>RSI+IPS</th>
<th>Ark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertices</td>
<td>520,105</td>
<td>465,788</td>
</tr>
<tr>
<td>Edges</td>
<td>1,034,228</td>
<td>934,326</td>
</tr>
<tr>
<td>Probes</td>
<td>2,073,988</td>
<td>4,042,521</td>
</tr>
<tr>
<td>Ingresses</td>
<td>38,787</td>
<td>31,110</td>
</tr>
<tr>
<td>Time</td>
<td>18h 33m</td>
<td>53h 48m</td>
</tr>
</tbody>
</table>
Current Status

• Implemented primitives on Ark:
  – Worked with CAIDA to debug, refine Ark interface
  – Integration into cohesive system
  – Operational experience gathering real topologies (amid load balancing, etc) using CAIDA’s topo-on-demand
• Have met year 1 milestones and deliverables
• Topology publication output:
  – **PAM2013**: “IPv6 Alias Resolution via Induced fragmentation”
  – **IMC2013**: “Speedtrap: Internet-scale IPv6 Alias Resolution”
  – **IMC2013**: “Internet Nameserver IPv4 and IPv6 Address Relationships”
  – **MILCOM2013**: “A Technique for Network Topology Deception”
Next Steps

• Probe whole Internet (rather than 50K subset)
• Begin multi-cycle probing using combined primitives
• Better quantify load savings and running time
• Begin gathering, analyzing, and reducing topologies to router and AS-level
• Tech transfer:
  – Working closely with CAIDA and Akamai
  – CAIDA will deploy an implementation of our primitives, beginning with IPv6 (to lower risk)
  – Planned activity for years 2 and 3
Contact Information

• Center for Measurement and Analysis of Network Data
  @NPS: http://www.cmand.org
• Contact:
  Robert Beverly
  Assistant Professor
  http://rbeverly.net/research
  rbeverly@nps.edu
  831-656-2132