CAIDA update

P. K. Clayfay, CAIDA
ISI/USC
Marina del Rey, CA
10 February 2017
CAIDA Update

• Data collection activities
  • Ongoing measurements
  • Data storage status
  • Data dissemination statistics
  • Recent publications

• Related other activities
  • New data infrastructure
  • Related research activities

• Open issues
  • Portal, New Data Types
Data Collection Infrastructures

- **Ark Platform (as of Sept 2016)**
  - 170 monitors in 59 countries
  - 74 IPv6-enabled
  - 124 Raspberry PIs

- **UCSD Network Telescope**
  - As of January 2017, captures more than 1TB of compressed traffic trace data per day.
  - 28 TB: last full month (Aug 2016)
  - 182 TB: 2015
  - 211 TB: YTD 2016 (as of 9/13/16)
  - 288 TB: last 12 months at NERSC (as of 9/13/16)
  - 703 TB: total archived at NERSC
<table>
<thead>
<tr>
<th>Datasets</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Toplogy Measurements w/ Skitter</td>
<td>0</td>
</tr>
<tr>
<td>OC48 Peering Point Traces</td>
<td>3</td>
</tr>
<tr>
<td>Backscatter</td>
<td>10 (4 rejected)</td>
</tr>
<tr>
<td>DDoS 2007 Attack Dataset</td>
<td>3 (1 rejected)</td>
</tr>
<tr>
<td>IPv4 2013 Census Dataset</td>
<td>3</td>
</tr>
<tr>
<td>IPv4 Routed /24 Topology</td>
<td>0</td>
</tr>
<tr>
<td>IPv4 Routed /24 DNS Names</td>
<td>0</td>
</tr>
<tr>
<td>IPv6 Topology</td>
<td>0</td>
</tr>
<tr>
<td>Internet Topology Data Kits (ITDK)</td>
<td>2 (1 withdrawn)</td>
</tr>
<tr>
<td>Patch Tuesday Dataset</td>
<td>3 (1 rejected)</td>
</tr>
<tr>
<td>Three Days of Conficker Dataset</td>
<td>4</td>
</tr>
<tr>
<td>Two-Days-in-2008 Telescope Dataset</td>
<td>3 (1 rejected)</td>
</tr>
<tr>
<td>UCSD Real-time Network Telescope Dataset</td>
<td>5 (1 rejected, 3 withdrawn)</td>
</tr>
<tr>
<td>UCSD Telescope Darknet Scanners Dataset</td>
<td>7 (1 rejected)</td>
</tr>
<tr>
<td>Witty Worm</td>
<td>2</td>
</tr>
</tbody>
</table>
New and Upcoming Data Sets

- (2) Macroscopic Internet Topology Data Kit (ITDK)
- IPv4 2013 Census Dataset
  http://www.caida.org/data/active/ipv4_2013_census_dataset.xml
  (available from IMPACT only)
- UCSD Network Telescope -- Darknet Scanners Dataset
  http://www.caida.org/data/passive/telescope-darknet-scanners_dataset.xml
  (available from IMPACT only)
- AS Border Mapping Dataset (coming soon)
- AS to Facilities Dataset (coming soon)
- Spoofer data

http://www.caida.org/data/publications/bydataset/index.xml#UCSD Network Telescope

<table>
<thead>
<tr>
<th>UCSD Network Telescope</th>
<th>102</th>
</tr>
</thead>
<tbody>
<tr>
<td>backscatter-2004-2005</td>
<td>8</td>
</tr>
<tr>
<td>backscatter-2006</td>
<td>4</td>
</tr>
<tr>
<td>backscatter-2007</td>
<td>8</td>
</tr>
<tr>
<td>backscatter-2008</td>
<td>17</td>
</tr>
<tr>
<td>backscatter-generic</td>
<td>4</td>
</tr>
<tr>
<td>backscatter-tics</td>
<td>3</td>
</tr>
<tr>
<td>code-red worm</td>
<td>7</td>
</tr>
<tr>
<td>code-red-generic</td>
<td>2</td>
</tr>
<tr>
<td>telescope-2days-2008</td>
<td>12</td>
</tr>
<tr>
<td>telescope-3days-conficker</td>
<td>14</td>
</tr>
<tr>
<td>telescope-educational</td>
<td>3</td>
</tr>
<tr>
<td>telescope-generic</td>
<td>7</td>
</tr>
<tr>
<td>telescope-patch-tuesday</td>
<td>2</td>
</tr>
<tr>
<td>telescope-real-time</td>
<td>5</td>
</tr>
<tr>
<td>witty worm (public)</td>
<td>1</td>
</tr>
<tr>
<td>witty worm (restricted)</td>
<td>17</td>
</tr>
<tr>
<td>witty-generic</td>
<td>6</td>
</tr>
</tbody>
</table>
Tools under consideration

- **Vela**: On-Demand Topology Measurement Service of CAIDA’s Ark infrastructure
  - Web interface [https://vela.caida.org/](https://vela.caida.org/)
  - Command-Line interface
Query Traces for IP Paths

Displays traceroute paths.

**Query**

Target Address/Prefix/AS/Country: 

Second Target for _neigh_ Query: 

Separate multiple targets with commas. 
Example: 1.2.3.4, 10.0.0.0/8, as1234, .sy

Start Date:  
End Date:  

Dates can be YYYY, YYYY-MM, or YYYY-MM-DD. End date is exclusive. 
Leave start/end (or both) blank for an open-ended range.

**Query Method:**

- dest  
- addr  
- neigh

**Target Position/Neighbor Separation:**  
Max Traces:  
Reverse Order

- **positive** position — hop distance relative to _beginning_ of trace
- **negative** position — hop distance relative to _end_ of trace
- **neighbor** separation — hop distance _between_ neighboring targets

**Vantage Point**

- By Name
- By Continent
- By Country
- By Org Type

Monitors with IPv6 have an asterisk next to their name.

Submit  
Reset
Tools: Henya

- **Henya**: Large-Scale Internet Topology Query System
  - Access via the Vela web interface [https://vela.caida.org/](https://vela.caida.org/)
  - 9 years of “Routed /24” trace routes
    - 47 billion traces in 20TB of files
    - growing yearly by 10 billion traces
  - 1 year of “Prefix Probing” trace routes
    - growing yearly by 9 billion traces

![Graph showing IPv4 data growth from 2007 to 2016]
Henya Topology Queries

• find occurrences of traceroute path elements
• ⟨targets⟩ = IP addresses, prefixes, ASes, or countries

• Queries:
  • traceroutes toward ⟨targets⟩
  • traceroutes containing one or more ⟨targets⟩

• Parameters:
  • measurement vantage points
  • data collection time periods
  • position of ⟨targets⟩ in path
  • hop distance between sets of ⟨targets⟩
Henya Query Complexity

- the most complex case:
  - traceroutes containing two or more \{targets\}
    - precisely: traceroutes containing some hop $h_1 \in \{targets_1\}$, $h_2 \in \{targets_2\}$, ...
  - example: traceroutes containing hops in both \{Germany\} and \{Japan\}

- harder:
  - traceroutes with hops in \{Germany or UK or France\} and hops in \{ATT or Level3 network\} and hops in \{Amsterdam Internet Exchange\}

find intersection of two sets
Vela and Henya Access Policies

- Currently accepting requests for accounts on Vela
- Currently accepting requests for early access to Henya and a subset of total topology dataset.
Restricted Dataset Requests

received/approved requests for restricted datasets

* This graph now includes all passive traces (including OC192). Previous graphs included only OC48 requests.

http://www.caida.org/data/about/
Users downloading public data

Number of users downloading public data

- Anonymized Internet traces
- UCSD Telescope
- Ark topology
- AS Relationships
- Ark AS links
- Skitter topology
- Other public topology
- Published data supplements

http://www.caida.org/data/about/
Users downloading restricted data

* This graph now includes all passive traces (including OC192). Previous graphs included only OC48 downloads.

http://www.caida.org/data/about/
Public data downloaded

Amount of public data downloaded

- Anonymized Internet traces
- UCSD Telescope
- Ark topology
- AS Relationships
- Ark AS links
- Skitter topology
- Other public data
- Published data supplements

http://www.caida.org/data/about/
Restricted data downloaded

Amount of restricted data downloaded

- Anonymized Internet traces
- UCSD Telescope
- Ark topology
- Witty
- Backscatter
- DDoS
- DNS lookups
- DNS root/gTLD RTT

• Drop in topology data in 2016 due to making topology data public

http://www.caida.org/data/about/
Recent Related R&D Activities

• DHS: Spoofing measurement (spoofer.caida.org)
• New DHS project: Science of Internet Security: Technology and Experimental Research (SISTER)
• NSF: Internet Outage Detection and Analysis (IODA) (ioda.caida.org)
• NSF: Internet congestion mapping system (beamer.caida.org)
Software Systems for Surveying Spoofing Susceptibility

- DHS S&T funded project that seeks to minimize Internet's susceptibility to spoofed DDoS attacks

- Goal: develop, build, and operate multiple open-source software tools to assess and report on the deployment of source address validation (SAV) best anti-spoofing practices.

- [https://spoofer.caida.org/] \(<— plz download now!\)

- Will share data through IMPACT
Software Systems for Surveying Spoofing Susceptibility

Recent Tests

<table>
<thead>
<tr>
<th>Session</th>
<th>Timestamp</th>
<th>Client IP</th>
<th>ASN</th>
<th>Country</th>
<th>NAT</th>
<th>Spoof Private</th>
<th>Spoof Routable</th>
<th>v4 Adjacency Spoofer</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>73442</td>
<td>2016-09-28 11:57:32</td>
<td>62.195.54.x</td>
<td>6830 (LG-UPC)</td>
<td></td>
<td>yes</td>
<td>rewritten</td>
<td>rewritten</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73440</td>
<td>2016-09-28 11:57:10</td>
<td>37.235.50.x</td>
<td>57169 (EDIS-AS-EU)</td>
<td></td>
<td>no</td>
<td>blocked</td>
<td>received 78</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73439</td>
<td>2016-09-28 11:57:07</td>
<td>84.59.214.x</td>
<td>3209 (VODANET)</td>
<td></td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73438</td>
<td>2016-09-28 11:51:56</td>
<td>95.90.233.x</td>
<td>31334 (KABELDEUTSCHLAND-AS)</td>
<td></td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73437</td>
<td>2016-09-28 11:49:27</td>
<td>91.14.132.x</td>
<td>3320 (DTAG)</td>
<td></td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73435</td>
<td>2016-09-28 11:47:31</td>
<td>79.237.172.x</td>
<td>3320 (DTAG)</td>
<td></td>
<td>yes</td>
<td>rewritten</td>
<td>rewritten</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73434</td>
<td>2016-09-28 11:43:39</td>
<td>94.214.191.x</td>
<td>9143 (ZIGGO)</td>
<td></td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73431</td>
<td>2016-09-28 11:36:16</td>
<td>70.196.30.x</td>
<td>22394 (CELLCO)</td>
<td>usa (United States)</td>
<td>yes</td>
<td>blocked</td>
<td>rewritten</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73429</td>
<td>2016-09-28 11:30:12</td>
<td>213.221.218.x</td>
<td>15600 (FINECOM)</td>
<td>che (Switzerland)</td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73426</td>
<td>2016-09-28 11:21:08</td>
<td>122.252.250.x</td>
<td>24180 (RAILTEL-AS-IN)</td>
<td>ind (India)</td>
<td>yes</td>
<td>unknown</td>
<td>unknown</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73424</td>
<td>2016-09-28 11:08:37</td>
<td>37.201.192.x</td>
<td>6930 (LG-UPC)</td>
<td>deu (Germany)</td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73423</td>
<td>2016-09-28 11:08:43</td>
<td>128.151.13.x</td>
<td>20 (UR)</td>
<td>usa (United States)</td>
<td>no</td>
<td>unknown</td>
<td>unknown</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73421</td>
<td>2016-09-28 11:06:25</td>
<td>91.154.254.x</td>
<td>719 (ELISA-AS)</td>
<td>fin (Finland)</td>
<td>no</td>
<td>unknown</td>
<td>unknown</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73420</td>
<td>2016-09-28 10:56:58</td>
<td>47.29.88.x</td>
<td>55836 (RELIANCEJO-IN)</td>
<td>ind (India)</td>
<td>yes</td>
<td>rewritten</td>
<td>rewritten</td>
<td>none</td>
<td>Full report</td>
</tr>
<tr>
<td>73419</td>
<td>2016-09-28 10:46:13</td>
<td>86.88.134.x</td>
<td>1136 (KPN)</td>
<td>nld (Netherlands)</td>
<td>yes</td>
<td>blocked</td>
<td>blocked</td>
<td>none</td>
<td>Full report</td>
</tr>
</tbody>
</table>

http://spoofer.caida.org/recent_tests.php
The video will explain to a general audience the dangers of IP spoofing.

Working towards a filtered tomorrow.

http://spoofer.caida.org

We will end the video with a requester help.
• Using the versatile Ark measurement platform, we will conduct measurements and analysis for documented explanations of structural and dynamic aspects of the Internet infrastructure relevant to cybersecurity vulnerabilities
  • Task 1: Support for Macroscopic Security and Stability Monitoring and Analysis
  • Task 2: Mapping Peering Interconnections at the Router Level
  • Task 3: Mapping Peering Interconnections at the Facility Level
  • Task 4: Measurements of TCP Behavior to Understand Security Vulnerabilities
  • Task 5: Identifying Grey Market IPv4 Address Transfers
  • Task 6: Internet Router-Level Topology Mapping on Demand
• **Task 1:**
  - IPv4 Prefix-Probing Dataset
    [http://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml](http://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml)

• **Task 2:**
  - AS Border Mapping Dataset (February 2017)

• **Task 3:**
  - AS to Facilities Mapping Dataset (February 2017)
  - AS to Facilities Mapping Dataset annotated w/ approach to interconnection (private peering with cross-connect, public peering, private interconnects over the public switch fabric, and remote peering) (February/March 2017)
  - Alias resolved Interconnection (router-level map) (April 2017)
  - Global facility-aware map of interconnection (May 2017)
Detection and analysis of large-scale Internet infrastructure outages (IODA)

• Developing methods to infer location and extent of outages

• Goals: (1) investigate and define strategies and methodologies to fuse diverse data sources to detect & characterize outages, (2) define and refine system requirements for continuous monitoring & (near) real-time analysis (3) testing & experimental deployment

• Part of a 3 year NSF-funded SATC project
Detection and analysis of large-scale Internet infrastructure outages (IODA)

Libyan outages: (a) visibility of Libyan IPv4 prefixes in BGP (RouteViews, RIPE NCC RIS);
(b) unsolicited traffic to UCSD telescope from Libya.
IODA After Four Years (Today)

- Live detection and monitoring

https://ioda.caida.org
• High-level system view
Mapping Interdomain Internet Congestion

• Developing methods to measure the location and extent of interdomain congestion

• **Goals (1)** system to monitor interdomain links and their congestion state, **(2)** near real-time “congestion heat map” of the Internet, **(3)** increase transparency, empirical grounding of debate

• Part of a 3 year NSF-funded project on topology+congestion
Measurement System

- **Topology measurement**
- **TSLP**
- **On-demand measurement**

**VP**

**AS relationship**

**BGP**

**IXP data**

**Link identification**

**WHOIS**

**Alias resolution**

**VP**

**topology data**

**Topology mapping**

**TSLP target selection**

**Measurement notification system**

**Links DB**

**Historical state**

**VP capability**

**Alerts DB**

**Alert system**

**Time series analysis**

**Data processing**

**Triggered meas scheduler**

**Triggered meas data**

**TSLP samples**

**Probing logic**

**Real-time querying**

**Visualization**

**Longitudinal views**

**Frontend**

**Backend System**
Mapping Interdomain Internet Congestion

Congestion seen between Comcast
Contact Information

PI: k claffy, CAIDA
kc@caida.org
http://www.caida.org/