Locality

a semi-formal flow dimension

John Gerth
Stanford University
Outline

• Dress for success
  – Semi-formal attire

• Locals only
  – Friends, acquaintances, and janitors
  – On the street where you live

• All along the (IPv4) watchtower
  – Where’d you say you were from?
  – Getting there is half the fun
What does “semi-formal” mean?

- **Formal attributes**
  - IP address, protocol, TTL, …
  - Required and universal

- **Semi-formal**
  - By convention – service port numbers
  - By context – TCP flags
  - By environment – VLAN tag
  - Derived or inferred from above
“Semi-formal” examples

• **SiLK/YAF**
  – INT/EXT address classification
  – Application Labeling

• **Argus**
  – Country Codes via Maxmind lookup
  – Flow status and state flags
Why have them?

• **Filtering**
  – Quickly remove extraneous data

• **Grouping**
  – Focus on flow semantics

• **Aggregate Behavior**
  – Inputs for modeling
Locality

• **Duality**
  – both internal and external components

• **Scope**
  – Most definitely defined by where you sit

• **Improve Hierarchy**
  – First-order formal definitions
  – Use context to extend with semi-formal levels
First-order Locality

• 0 : announcement
  – Broadcast (normally x.y.z.255)
  – Multicast ( 224.0.0.0/4 )

• 1 : conversational
  – All unicast IP traffic
Extended Internal Locality

• 2 : Enterprise conversational traffic
  – All IP ranges owned by enterprise
  – Includes any RFC 1918 ranges
    • 10.0.0.0/8
    • 172.16.0.0/12
    • 192.168.0.0/16
  – And autoconfiguration
    • 169.254.0.0/16
Organizational Locality

• **3 or higher: enterprise sub-domains**
  – Likely limited by location of flow collection
  – Could also have multiple levels
  – Could be derived from other value
    • Subnet number
    • VLAN tag
    • Internal department/operating unit designation
Implementation

• **Goals**
  – Locality defined by IP address
  – First class dimension for filter and aggregation
  – Handle partial sub-allocation
  – Real-time annotation of flow data

• **Solution**
  – ASCII config file
  – Generate binary table indexed by IP/24 prefix
Example: Stanford CS

• Enterprise Entries

38.114.142.0/23  32  2
128.12.0.0/16   32  2
171.64.0.0/14   32  2
204.152.100.0/22 32  2
172.16.0.0/12   32  2
...

• Departmental Sub-allocation Override

171.67.76.0/23  32  3816
172.27.76.0/23  32  3816
...

FloCon 2015
Extended External Hierarchy

• **Motivation**
  – Better granularity for classifying traffic
  – Mitigate games of Whac-a-Mole in the hairball

• **Hierarchical Dimension Choices**
  (could choose more than one)
  – Subnet, e.g. CIDR/16
  – Geolocation data
  – Autonomous System Number (ASN)
Autonomous Systems

• Formal leaf nodes of the internet
  – Complement geography with “netography”
  – Aggregation point for enterprises

• Drive traffic at the “wholesale” level
  – ASN fuels the BGP tables

• ASNs are highly correlated to ISPs
  – Where most abuse complaints need to go
Mapping IP ranges to ASNs
(rather than monitoring BGP in real-time)

• Maxmind (monthly)
  – http://dev.maxmind.com/geoip/legacy/geolite/

• CAIDA (daily)
  – http://www.caida.org/data/routing/routeviews-prefix2as.xml

• Team Cymru (updates every 4 hours)
  – http://www.team-cymru.org/Services/ip-to-asn.html

• Routeviews (hourly)
  – http://www.routeviews.org/
Locality for Stanford EE/CS

- **Observation point**
  - Layer 2 entry point switches of three buildings
- **Topology**
  - Four dozen VLANs shared across buildings
- **Locality definition**
  - 0, 1, 2, VLAN
- **Flow storage**
  - SQL-like relational DB
Sample Queries

• Monitor overall locality distribution

```
h "select flows:count i, log_appbyte:10 xlog sum t_ab  by  locality:3 & loc, p:proto from flow where proto<>1"
locality p | flows    log_appbyte
-----------| -------------------
0        17| 2597085   9.2
1        6 | 1711643  12.6
1        17| 3140121  10.6
2        6 | 3885930  11.8
2        17| 13417251 10.6
3        6 | 4313177  12.8
3        17| 11861066 11.3
```
Sample Queries

- **Top IPs after removing service ASNs**

  "Top Remote except Google (15169) + Amazon (16509)"

<table>
<thead>
<tr>
<th>asn</th>
<th>ripn</th>
<th>nlip</th>
<th>tot</th>
<th>ix</th>
<th>begin</th>
<th>recent</th>
</tr>
</thead>
<tbody>
<tr>
<td>46664</td>
<td>199.168.136.95</td>
<td>832</td>
<td>344328</td>
<td>0.555</td>
<td>20:47</td>
<td>23:59</td>
</tr>
<tr>
<td>31042</td>
<td>94.189.239.232</td>
<td>519</td>
<td>191031</td>
<td>0.555</td>
<td>10:29</td>
<td>18:59</td>
</tr>
<tr>
<td>21581</td>
<td>108.161.147.110</td>
<td>47</td>
<td>183337</td>
<td>0.376</td>
<td>00:00</td>
<td>23:59</td>
</tr>
<tr>
<td>36024</td>
<td>74.50.54.108</td>
<td>45</td>
<td>155905</td>
<td>0.415</td>
<td>00:00</td>
<td>23:59</td>
</tr>
<tr>
<td>4134</td>
<td>222.95.211.39</td>
<td>833</td>
<td>124722</td>
<td>0.0851</td>
<td>01:27</td>
<td>12:29</td>
</tr>
<tr>
<td>4134</td>
<td>115.231.222.176</td>
<td>149</td>
<td>93499</td>
<td>0.241</td>
<td>11:28</td>
<td>23:59</td>
</tr>
<tr>
<td>3842</td>
<td>167.88.124.163</td>
<td>1</td>
<td>86332</td>
<td>-0.000533</td>
<td>00:00</td>
<td>23:59</td>
</tr>
<tr>
<td>32934</td>
<td>185.60.216.7</td>
<td>739</td>
<td>84821</td>
<td>-0.189</td>
<td>00:00</td>
<td>23:59</td>
</tr>
<tr>
<td>12876</td>
<td>62.210.180.31</td>
<td>86</td>
<td>81358</td>
<td>0.253</td>
<td>00:00</td>
<td>23:51</td>
</tr>
<tr>
<td>4134</td>
<td>117.89.17.200</td>
<td>733</td>
<td>78038</td>
<td>0.0784</td>
<td>12:36</td>
<td>16:25</td>
</tr>
</tbody>
</table>
Sample Queries

- Chase internal spam source

```
h \"select f:count i by vlan \ from flow where d_ip=171.64.y.z, d_port=25, loc>1\" vlan| f ----|----- 3803| 57747 3864| 1451

# Now 'pivot' on vlan
h \"select f:count i by ips s_ip from flow where d_ip=171.64.y.z,d_port=25,vlan=3803\" s_ip | f ----------------|----- 172.24.15.162| 185 172.24.15.164| 22745 172.24.15.175| 30287 172.24.15.178| 135 172.24.15.185| 3205 172.24.15.190| 63 172.24.15.9 | 1127
```

FloCon 2015

Stanford Computer Science
Future Work

• True real-time updates to locality
  – Internal via DNS + DHCP updates
  – External via BGP monitor
• Extending external hierarchy
  – Country code
  – Additional Geolocation
• IPv6

FloCon 2015
Summary

• Every IP has an ASN
  – Either the enterprise ASN – or the remote ASN when locality is 1
  – srcASN = ASmap[srcIP]; dstASN = ASmap[dstIP]

• Every flow has a locality
  \( \text{Let } \text{uni}=\{? \text{ unicast dstIP}\}; \text{ then } \text{locality}= \text{uni} *( \text{uni} + (\text{srcASN} == \text{dstASN}) ) \)
  – 0: non-unicast
  – 1: unicast from outside enterprise
  – 2: enterprise unicast outside observation point
  \( \text{ optionally } \)
  – 3+: additional granularity inside organizational unit