Large-Scale Geolocation for NetFlow

Pavel Čeleda, Petr Velan, Martin Rábek
Rick Hofstede, Aiko Pras
{celeda|velan|xrabek1}@ics.muni.cz, {r.j.hofstede|a.pras}@utwente.nl

IFIP/IEEE IM 2013, 27-31 May 2013, Ghent, Belgium
Part I

Introduction
Motivation and R&D Goals – I

SURFmap - a Network Monitoring Tool Based on the Google Maps API.
Motivation and R&D Goals – II

How flow-based geolocation can be performed in a large-scale?

- exporter-based approach,
- collector-based approach.

How can we benefit from geolocation data in flow records?

- traffic engineering,
- traffic profiling,
- anomaly detection.
Part II

Architecture
Exporter-Based Geolocation

Packets → Input → Flow cache → Export → NetFlow v9

exporter filter plugin for IP address geolocation,
NetFlow v9 template mapping – GEO data to AS fields
SRC_AS=*SRC_GEO, DST_AS=*DST_GEO

AS mapping → transparent to any flow collector.
Exporter-Based Geolocation

Packets → Input → Flow cache → Export → NetFlow v9

Flows

GeoPlugin

Flows → Geolocated flows

exporter filter plugin for IP address geolocation, NetFlow v9 template mapping – GEO data to AS fields
SRC_AS=*SRC_GEO, DST_AS=*DST_GEO

AS mapping → transparent to any flow collector.
Exporter-Based Geolocation

- exporter filter plugin for IP address geolocation,
- NetFlow v9 template mapping – GEO data to AS fields
  `SRC_AS=*SRC_GEO`, `DST_AS=*DST_GEO`,
- AS mapping $\rightarrow$ transparent to any flow collector.
MaxMind GeoLite Country Database

- MaxMind GeoLite – free off-line country database,
- C-API for IPv4/IPv6 geolocation.


IPv4 IPv6 Queries/s (x 10^6)

Standard Memory cache Check cache MMAP cache

Pavel Čeleda

Large-Scale Geolocation for NetFlow
Collector-Based Geolocation

Data collection
NetFlow
v5, v9
nfcapd
Geolocation

- patch for NFDUMP and NfSen toolset,
- native geolocation support for any NetFlow v5/v9, IPFIX data.
patch for NFDUMP and NfSen toolset,
native geolocation support for any NetFlow v5/v9, IPFIX data.
patch for NFDUMP and NfSen toolset,

native geolocation support for any NetFlow v5/v9, IPFIX data.
Flow Record:

Flags        = 0x06 Unsampled
size         = 80
first        = 1348387461 [2012-09-23 10:04:21]
last         = 1348387462 [2012-09-23 10:04:22]
msec_first   = 890
msec_last    = 100
src addr     = 23.63.79.144
dst addr     = 147.251.170.165
src port     = 80
dst port     = 57046
tcp flags    = 0x1a .AP.S.
proto        = 6
(in)packets  = 4
(in)bytes    = 936
input        = 5
src as       = 20940
dst as       = 2852
in src mac   = 00:0e:38:5e:30:c0
out dst mac   = 00:1e:be:8b:26:c0
src ctry     = 840 ... ISO 3166-1 country code - US
dst ctry     = 203 ... ISO 3166-1 country code - CZ
NFDUMP Flow Listing

a) numeric code – %scc %dcc

<table>
<thead>
<tr>
<th>Proto</th>
<th>Src IP Addr:Port</th>
<th>Dst IP Addr:Port</th>
<th>Src Ctry</th>
<th>Dst Ctry</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP</td>
<td>194.228.29.173:0</td>
<td>147.251.48.205:3.13</td>
<td>203</td>
<td>203</td>
</tr>
<tr>
<td>UDP</td>
<td>151.40.40.243:15833</td>
<td>147.251.79.246:49159</td>
<td>380</td>
<td>203</td>
</tr>
<tr>
<td>TCP</td>
<td>157.55.235.165:40040</td>
<td>147.251.215.10:49464</td>
<td>840</td>
<td>203</td>
</tr>
<tr>
<td>UDP</td>
<td>147.251.170.77:59408</td>
<td>89.79.20.120:18973</td>
<td>203</td>
<td>616</td>
</tr>
</tbody>
</table>

b) alpha-2 code – %sccan %dccan

<table>
<thead>
<tr>
<th>Proto</th>
<th>Src IP Addr:Port</th>
<th>Dst IP Addr:Port</th>
<th>Src Ctry</th>
<th>Dst Ctry</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP</td>
<td>194.228.29.173:0</td>
<td>147.251.48.205:3.13</td>
<td>CZ</td>
<td>CZ</td>
</tr>
<tr>
<td>UDP</td>
<td>151.40.40.243:15833</td>
<td>147.251.79.246:49159</td>
<td>IT</td>
<td>CZ</td>
</tr>
<tr>
<td>TCP</td>
<td>157.55.235.165:40040</td>
<td>147.251.215.10:49464</td>
<td>US</td>
<td>CZ</td>
</tr>
<tr>
<td>UDP</td>
<td>147.251.170.77:59408</td>
<td>89.79.20.120:18973</td>
<td>CZ</td>
<td>PL</td>
</tr>
</tbody>
</table>

Usage example

nfdump -M /data/nfsen/profiles-data/live/p3000:p3001 \
   -r 2012/09/23/nfcapd.201209231005 \
   -o 'fmt:%pr %sap -> %dap %sccan %dccan' -m -c 20
**Geofiltering**

- Country filter syntax is similar to other NFDUMP filters syntax: ctry [comp] <num>,
- Country can be compared to a list (red-black tree) of country codes, syntax: ctry in [ <ctrylist> ],
- Filters are often used for traffic profiling in NfSen.

**Usage example**

```
nfdump -M /data/nfsen/profiles-data/live/p3000:p3001 \  
  -r 2012/09/23/nfcapd.201209232035 -c 5 \  
  'src ctry 203 and not dst ctry in [ 203 840 166 ]'
```
NfSen Geoprofiling

: Screenshot of collector-based geolocation prototype.
Part III

Use Case I – Traffic Profiling
ICMP traffic.
Geolocated and Non-geolocated ICMP Traffic – II

: Geolocated ICMP traffic.
Distribution of HTTPS Traffic over Countries – I

HTTPS flows/s.

IN
OUT
US
CZ
Other

Flows/s

US
CZ
Other

Pavel Čeleda
Large-Scale Geolocation for NetFlow
Part IV

Use Case II – Anomaly Detection
Bad Neighboring Countries

Incoming TCP SYN-only flows.
UDP DoS attack from infected Linux machine.
Part V

Conclusion
Conclusion

Summary

- country-level information in flow data,
- native geolocation support for NfSen/NFDUMP,
- pilot geo-prototype deployment at MU – CESNET link.

Future Work

- IPFIX-compliant prototype for exporter-based geolocation,
- ipfixcol – AS and GEO support implementation,
- AS + GEO data for traffic profiling and anomaly detection.
Thank You For Your Attention!

UNIVERSITY OF TWENTE.

P. Čeleda, P. Velan, M. Rábek
{celeda|velan|rabe}@ics.muni.cz

R. Hofstede, A. Pras
{r.j.hofstede|a.pras}@utwente.nl

Geolocation Toolset
http://www.muni.cz/research/publications/1090804