Role Model Transformations for Flow Analysis in Cyberdefense

John Gerth

Stanford University

STANFORD
COMPUTER SCIENCE

FloCon 2016

Metrics and Analytics

- 2012 Packet Dynamics
  - Leveraging protocols and human factors
- 2014 Producer/Consumer Ratio
  - Characterizing data flow across scales
- 2015 ASNs and Flow Locality
  - Enriching “netography” in flow
- 2016 Orientation and Degree
  - The view from where you sit

Classic Flow Orientation and Role

- Orientation
  - – Packet Src / Dst Addresses
- Role
  - – Client/Server
    - By convention – first packet seen
    - By context – "well-known" port numbers
    - Derived or inferred from above

STANFORD
COMPUTER SCIENCE

Cyberdefense Tasks

- Situational Awareness
  - Summarize “the view from here”
  - Which ways are data flowing?
  - What roles are evident?
- Incident Response / Forensics
  - Focus on flow sequences and semantics
  - Who have suspect machines been talking to?
  - Can we pivot our viewpoint?

Identifying Roles

- Across Control/Data/Management Planes
    - Client/Server
    - Master/Slave
    - Peer-to-Peer
    - Pub/Sub
- Behavioral Domains
    - Not just Src/Dst -- Location of actors is critical
    - Producer/Consumer
    - Importer/Exporter
    - It’s mostly “us versus them”
        - North/South
    - Although “them” may be internal
        - East/West

Locality Layers

- Every IP has a locality

  Either the enterprise ASN – or the remote ASN
  srcASN = ASmap[srcIP]; dstASN = ASmap[dstIP]

- Every flow has a locality

  (**Let** uni=:{? unicast dstIP}; **then** locality:= uni *(uni + (srcASN == dstASN))

  0: non-unicast

  1: unicast from outside enterprise

  2: enterprise unicast from outside observation point

  (optionally)

  3+: additional east/west granularity inside organizational units

Calculating Flow Locality

- Every Flow acquires a locality from its IPs
  - Compare locality value of SrcIP and DstIP
  - **Lower** value is the value for the flow
- Example: DNS lookup via Google
  - SrcIP = 172.17.1.34 (locality = 2)
  - DstIP = 8.8.8.8 (locality = 1)
  - Flow locality = 1

Flow Degree

- Summarize Orientation and Data Exchange

  Capture the local perspective on communication
- Sign gives direction of first packet

  +: Local IP (outbound)

  -: Remote IP (inbound)
- Exchange depth (decreasing order)

  3: Payload exchange in both directions

  2: Packet exchange in both directions

  1: One-way packets

  0: Nonsensical (for bi-directional flows)

STANFORD
COMPUTER SCIENCE

Flow Degree Values

3: Local client exchanged payloads

2: Local client exchanged packets

1: Local client sent ignored packets

0: Nonsensical

-1: Local server ignored packets

-2: Local server exchanged packets

-3: Local server exchanged payloads

Degrees – Good, Bad, and Ugly

3 : Local client exchanged payloads with remote server
- Lookup at Google Public DNS
- Data exfiltration via DNS to Ukranian domain

2 : Local client exchanged packets with remote server
- Normal exchange of ACKs with Amazon web server
- Heartbeat sent to Dutch C&C server

1: Local client sent ignored packets to remote address
- Connect fails to offline webserver at non-profit
- Compromised laptop in marketing scanning DOE lab

0: Nonsensical
- DDOS backscatter (SYN/ACK, NTP spoof, …)

-1: Local server ignored packets from remote address
- Host firewall silently drops Brazilian RDP troll
- Compromised desktop receives "port-knock" packet from asian IP

-2: Local server exchanged packets with remote client
- FIN/ACK during web server TCP session teardown
- ACK sent during DDOS SYN flood

-3: Local server exchanged payloads with remote client
- Mail server accepts inbound Greek e-mail for local user
- Web server compromised by SQL injection from Romanian IP

Transforms

- Definitions

  F.direction =: SIGN(F.degree)
  F.locindex =: (1 = F.dir ? 0: 1)

- Addresses

  Lip=: (Sip,Dip)[F.locindex]
  Rip=: (Sip,Dip)[~F.locindex]

  Sip=: (Lip,Rip)[F.locindex]
  Dip=: (Lip,Rip)[~F.locindex]

- Metrics
  - Producer/Consumer Ratio (1.0 to -1.0)
    - PCR=: (Spayload-Dpayload)/(Spayload+Dpayload)
  - Export/Import Ratio (1.0 to -1.0)
    - XIR=: (Lpayload-Rpayload)/Lpayload+Rpayload)
  - Relationship
    - XIR = F.direction * PCR

Implementation at Stanford EE/CS

- Observation point
  - – Layer 2 entry point switches of three buildings
  - – Argus sensor creating bi-directional flows
- Topology
  - – Four dozen VLANs shared across buildings
- Locality definition
  - – 0, 1, 2, VLANid

Classic Flow Storage

- Archives for batch analysis
  - – Flat flow files organized by sensor
  - – Sequentially time-sequenced

- Relational DB for interactive queries
  - – Tables partitioned by date/time
  - – Indexed by Src IP, Dst IP, Dst Port

Enhanced Database Organizaton

- Addresses
  - Store both Src/Dst and Lcl/Rmt
  - Rmt ASN
- Degree + Locality
  - For each flow store Degree + Locality
- Indexing
  - Lcl/Rmt IPs
  - Dst port, ASN, VLAN

STANFORD
COMPUTER SCIENCE

Situational Awareness Queries

- Aggregate traffic by date for last 96 hours

```
q)select f:count i, count distinct l_ipn, count distinct r_ipn,  xir:avg pcr*signum role, sum t_ab by date  from
flow where date within 2015.06.23 2015.06.26
date      | f       l_ipn r_ipn xir   t_ab
----------| ---------------------------------------
2015.06.23| 9241197 6057  69320 0.22  3049916808392
2015.06.24| 7833157 6096  63296 0.277  495980015533
2015.06.25| 8083707 5976  59831 0.279  360244608240
2015.06.26| 8365180 6038  56958 0.28  1988082088281
```

- Today’s traffic by flow degree

```
q)select f:count i, count distinct l_ipn, count distinct r_ipn,  xir:avg pcr*signum role, sum t_ab by  deg from flow
where date=2015.06.26
deg | f       l_ipn r_ipn xir    t_ab
----| ---------------------------------------
-3  | 719903  616   18933 0.239  939859443721
-2  | 113111  611   9625  -0.235   2070257797
-1  | 179180  4916  11756 -0.61      40699133
0   | 23377   188   943   0          14913012
1   | 3016053 3516  14408 0.961   28581829274
2   | 775767  937   16088 0.228   53623522525
3   | 3537789 1395  22515 -0.217 963891422819
```

STANFORD
COMPUTER SCIENCE

Situational Awareness Queries

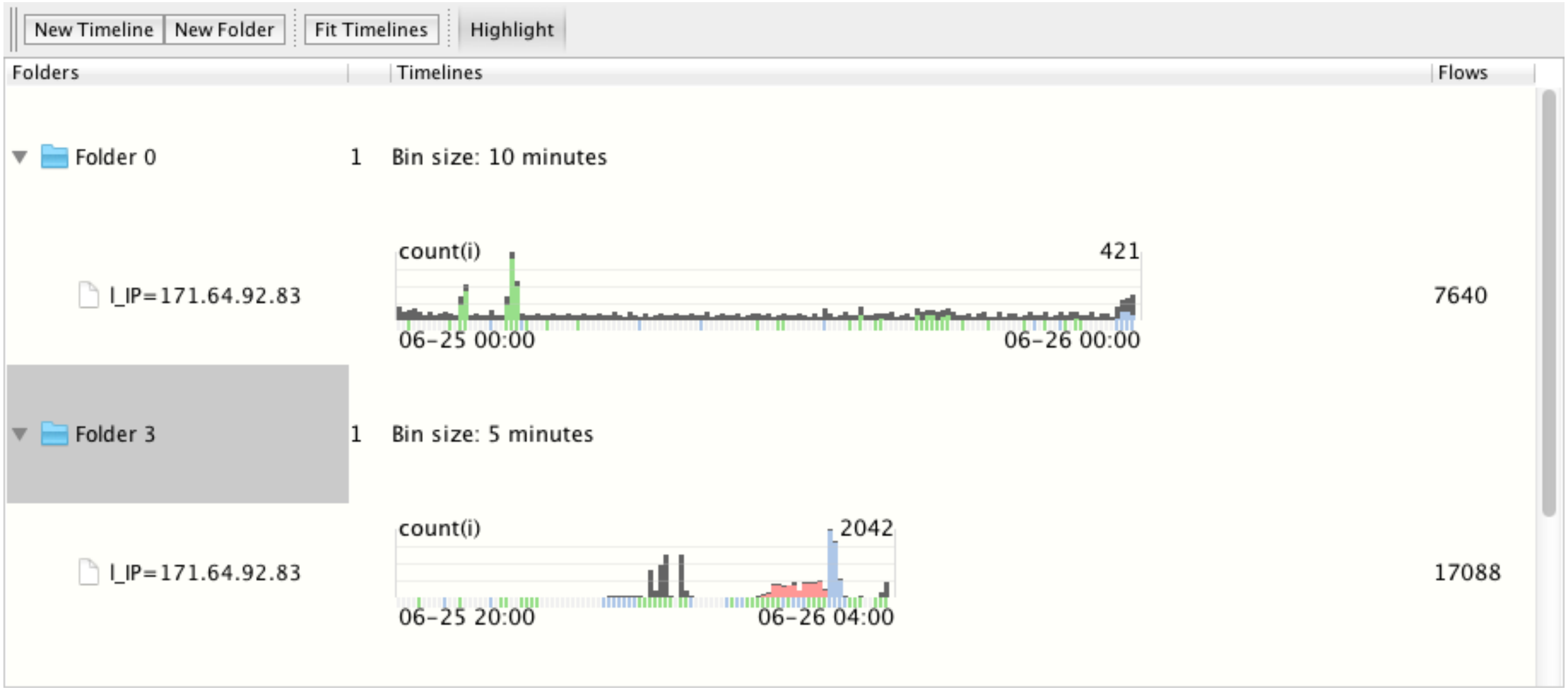
- Show dataflow for top remote hosts

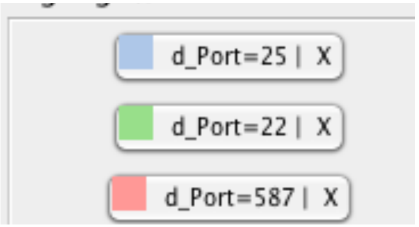
```
"Top Remote (excluding Google, Amazon, Yahoo)"
asn   ripn            nlip tot     xi       begin recent
--------------------------------------------------------
45899 113.160.41.218  1    1126982 0.974    00:00 18:04
36375 141.212.109.57  2    313645  -0.369   00:00 18:04
27385 64.39.103.75    1    218670  0.625    04:59 16:14
21581 108.161.147.110 47   143908  0.318    00:00 18:04
24940 136.243.74.81   2    135902  -0.999   00:00 18:04
```

Incident Handling

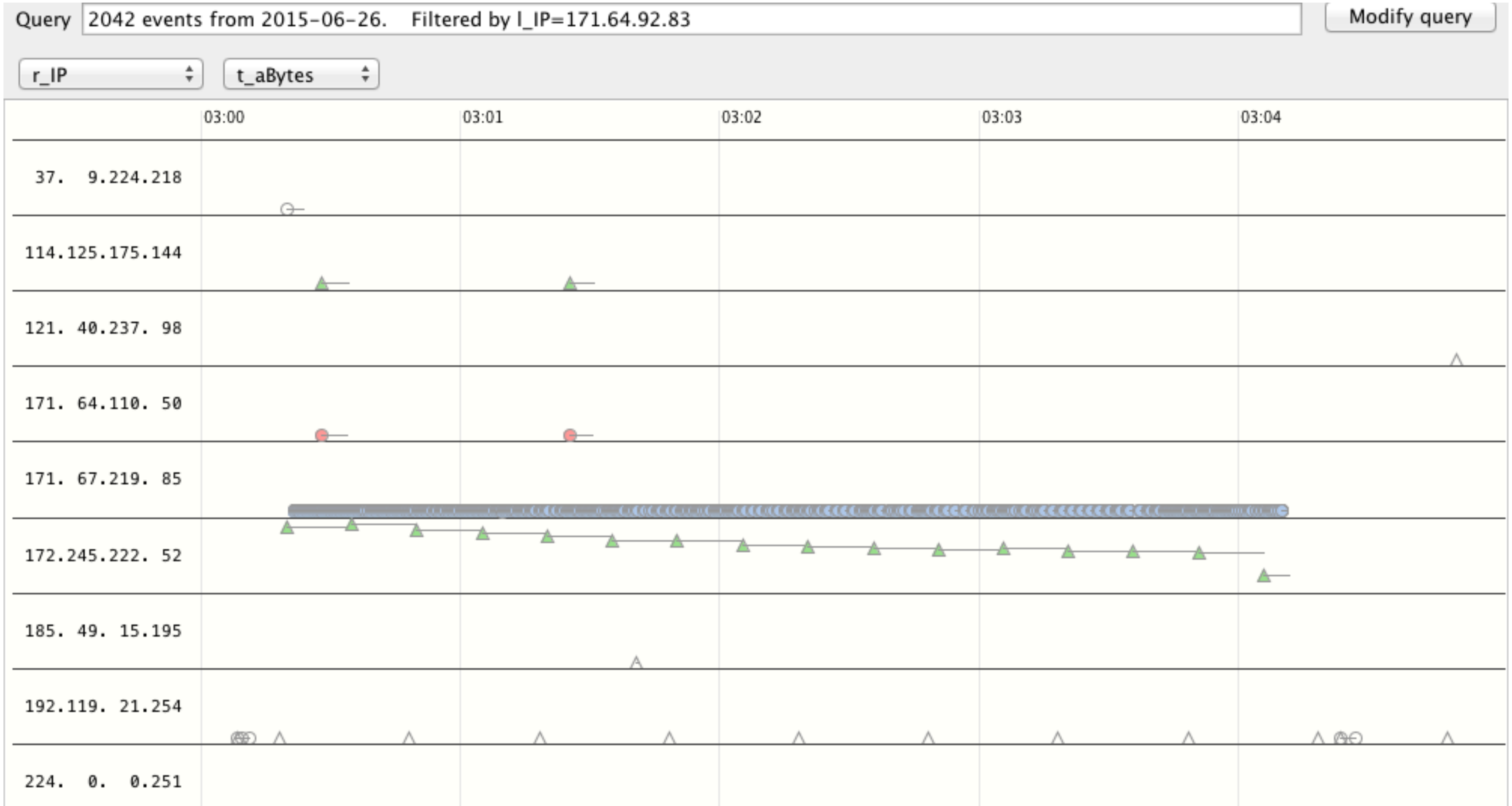
```
q)select f:count i by date from flow where date within 2015.06.21 2015.06.26,l_ipn=ipi `171.64.92.83
date      | f
----------| ------
2015.06.21| 7716
2015.06.22| 8646
2015.06.23| 7721
2015.06.24| 10121
2015.06.25| 7640
2015.06.26| 104374

q)select f:count i by loc,deg from flow where date=2015.06.26,l_ipn=ipi `171.64.92.83,(abs loc)in 1 2 3h
loc deg | f
--------| -----
1   -3  | 6255
1   -2  | 245
1   -1  | 247
1   0   | 94
1   1   | 480
1   2   | 65
1   3   | 501
2   -2  | 1
2   -1  | 38
2   1   | 1383
2   2   | 247
2   3   | 82846
```

Intrusion Timelines





Intrusion Detail



Query 2042 events from 2015-06-26. Filtered by l_IP=171.64.92.83

r_IP

t_aBytes

Summary

- Enabling Additional Perspectives
  - North/South and East/West
  - Export/Import, Depth, and Pivot
- Future work
  - Developing robust role signatures
  - Turning flow sequences into behaviors
  - Models of expected roles and behaviors

STANFORD
COMPUTER SCIENCE