A Rendezvous-based Paradigm for Analysis of Solicited and Unsolicited Traffic

DUST 2012
May 15, 2012

David Plonka
&
Paul Barford
{plonka,pb}@cs.wisc.edu
Outline

• Rendezvous-based Traffic Analysis
  – What is it? Why use it?
  – a DNS rendezvous case study involving office and residential “solicited” traffic

• Darkspace Rendezvous Mechanisms
  – unsolicited and passively solicited traffic

• TreeTop
  – a DNS rendezvous-based analysis tool
    [Plonka & Barford, IMC 2009, SATIN 2011, work in progress]
    – flow export with rendezvous annotations
    – IPv6 performance by service names
Rendezvous-based Traffic Analysis?

- Traffic classification and analysis has focussed on target traffic features (IP headers, DPI, etc.)
- However, Internet hosts learn IP addresses by some *rendezvous* mechanism, e.g.:
  - By static configuration (IP addrs in config files)
  - The Doman Name System (DNS)
  - Application-specific mechanisms (URLs, p2p)
- Inform traffic analysis by considering, “How does this host know this IP address?” rather than simply, “With what IP address did this host interact?”
Why Focus on Rendezvous?

*rendezvous*, meaning hosts and services “present themselves”

- For standard protocols, rendezvous information is not private and is of low-volume
  - Separate and separable from private payloads
  - Can be monitored in situations where target traffic is high-volume, sampled, or encrypted

- Rendezvous info can indicate when other analysis or classification techniques are effective and not
  - e.g., port-based classification

[Kim, et al., 2008] [Plonka & Barford, 2011]
Rendezvous-based Traffic Classification

rendezvous, meaning “present yourselves”

• **Hypothesis**: We can inform and improve traffic classification by considering, “How does this host know that peer IP address?”

• **DNS**: Internet hosts regularly use the DNS to find remote IP addresses of the hosts with which they might interact.
  – It is an *easily separable* standard, “clear text” protocol.
DNS Rendezvous: (1) Query
DNS Rendezvous: (2) Response

A 192.0.2.1
DNS Rendezvous: (3) Outbound
DNS Rendezvous: (4) Inbound
Traffic Observation Points

Internet host

DNS server

client
Traffic Observation Points
Traffic Observation Points

DPI, low-volume
Traffic Observation Points

SPI, high-volume

DPI, low-volume
Characteristics of Data Sets

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Date</th>
<th>Day</th>
<th>Duration</th>
<th>Clients</th>
<th>Unique NOERROR FQDNs</th>
<th>DNS Reply Pkts</th>
<th>Average DNS Reply Utilization</th>
<th>Average Wide-Area Outbound / Inbound Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>2009-04-17</td>
<td>Fri</td>
<td>24h</td>
<td>614</td>
<td>19.4 K</td>
<td>560 K</td>
<td>12.2 Kbps</td>
<td>753 Kbps / 5.66 Mbps</td>
</tr>
<tr>
<td>Residential</td>
<td>2009-04-17</td>
<td>Fri</td>
<td>24h</td>
<td>9,819</td>
<td>(5,344)</td>
<td>(143 K)</td>
<td>15.7 M</td>
<td>360 Kbps / 244 Mbps / 276 Mbps</td>
</tr>
</tbody>
</table>

![Graph showing local time of day vs. bits per second, with separate lines for outbound and inbound rates.](image1)

![Graph showing local time of day vs. bits per second, with separate lines for outbound and inbound rates.](image2)
Target Traffic Classification: Port-based method
Residential: Domain Popularity
Office Target Traffic Classification: “named” and “unnamed”
Residential Target Traffic Classification: “named” and “unnamed”
Residential Target Traffic Classification: “named” by popular domains
Host Profiling and Reputation based on Rendezvous Information

Venn Diagram:
- **Torrent**
  - Within Game: 13
  - Outside Game: 303
  - Total: 417 (7.5%)
- **Talk**
  - Within Game: 20
  - Outside Game: 626
  - Total: 717 (12.8%)
- **Game**
  - Total: 226 (4.0%)
  - Total intersections: 58
  - Total outside intersections: 303 + 626 + 226 - 58 = 1097
Residential Hosts Classification by P2P Host Profile (1 day)

- Torrent: 417 (7.5%) - 303
- Talk: 717 (12.8%) - 626
- Game: 226 (4.0%) - 150

Intersection counts: 58, 13, 20
“unnamed” Target Traffic by P2P Profile

Torrent (67.1%) 37.8%

2.6%

4.5%

5.9%

22.2%

Talk (20.0%) 7.0%

Game (34.9%) 2.3%
## Results Summary: Traffic Classified (% bytes)

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Port-known</th>
<th>DNS-named and Port-known</th>
<th>DNS-named</th>
<th>DNS-named and DNS-Profiled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office Out</td>
<td>93.9%</td>
<td>80.5%</td>
<td>81.8%</td>
<td>91.9%</td>
</tr>
<tr>
<td>Office In</td>
<td>96.6%</td>
<td>91.8%</td>
<td>93.2%</td>
<td>95.4%</td>
</tr>
<tr>
<td>Residential Out</td>
<td>18.6%</td>
<td>6.2%</td>
<td>6.7%</td>
<td>83.5%</td>
</tr>
<tr>
<td>Residential In</td>
<td>76.9%</td>
<td>58.3%</td>
<td>67.9%</td>
<td>88.2%</td>
</tr>
</tbody>
</table>
Rendezvous in Darkspace/Grayspace?

- **Darkspace and Unsolicited**: a host uses some technique to choose remote/peer IP addresses
  - **Algorithm**, e.g., scanning a contiguous set of IP addresses in series, choosing IP addresses at random
  - **Bug**, e.g. D-link products connect to 45.52.84.48, the 7-bit string “-4T0”, believed to be a stray value left in an uninitialized 32-bit integer meant to store an SMTP server's IP address [Yegneswaran, Barford, Plonka, 2004]
  - **Misconfiguration** or stale configuration, e.g., SNMP traps to various 45/8 addresses from Interop events
  - IP prefixes become **encumbered** by legacy roles
TreeTop:
Rendezvous-annotated Flow Export
TreeTop: radix tries and domain trees
[3 private slides redacted]
Discussion

● In what circumstances can we trust rendezvous information for traffic classification or host profiling/reputation?

● Tap rendezvous methods other than the DNS; e.g., application-specific methods (WWW, P2P); are they discoverable, separable and clear?

● Should we alter or invent rendezvous protocols to better inform classification and packet treatment?

● Is rendezvous a useful unifying analysis concept?
A Rendezvous-based Paradigm for Analysis of Solicited and Unsolicited Traffic

FIN

David Plonka
&
Paul Barford
{plonka, pb}@cs.wisc.edu