Assessing Performance of Internet Services on IPv6

The Eighteenth IEEE Symposium on Computers and Communications (ISCC '13) July 10, 2013



David Plonka

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

Motivation

- The era of IPv6 is obviously upon us
 - Many popular services operate on Ipv6
 - Yet early IPv6 experiences are unpredictable
- Good end-to-end performance is challenging
 - Deployment and operation in parallel with IPv4
 - New network configuration/monitoring tasks
 - Variety of client host configurations:
 - IPv4 only, IPv6 only, *Dual-Stack*
- Which protocol does a client choose?
 Which should it choose and why?

Approach

- Rendezvous-based Traffic Analysis

 What is it? Why use it?
- Implementation: TreeTop
 - a DNS rendezvous-based analysis tool
 [Plonka & Barford, IMC 2008, SATIN 2011]
 - flow export with rendezvous annotations
- Sample Applications:
 - Aggregate traffic measurement by service
 - Classification of encrypted traffic (e.g., HTTPS)
- Passive performance measurement of services on IPv6 versus IPv4

Rendezvous-based Traffic Analysis

- Traffic classification and analysis has focused 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 Domain Name System (DNS)
 - Application-specific mechanisms (URLs, p2p)
- We 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: how hosts "present themselves"

- For most popular protocols, rendezvous information is not secret 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 be detected a priori: ideal for caching and high-speed operation
- Rendezvous info can indicate when other analysis or classification techniques are effective and when they're not
 - e.g., bolstering *port-based classification* [Kim, et al., 2008] [Plonka & Barford, 2011]

Why Focus on IPv6 Rendezvous?

Rendezvous: how hosts "present themselves"

- We argue that the rendezvous mechanism is practically the only thing in common between accessing a service by both IPv4 and IPv6
 - The two protocol versions use entirely separate IP address namespaces (or numberspaces) that would otherwise require insider knowledge of the service implementation

• Here we will focus on the DNS, *i.e.,* similar A and AAAA queries



Traffic Observation Points





Rendezvous-annotated Flow Export

TreeTop uses two annotation approaches for flow source and destination addresses:

- *Direct:* TreeTop discovers that the given client end-host knows a remote IP address by a domain name from a prior DNS A or AAAA query
- Consensus: we infer, by shared consensus of other client end-hosts, that the hosts could have used the DNS to similarly resolve the peer's name. Name sampling is performed to clarify otherwise ambiguous names.



TreeTop enhanced with nmsg support

We select **nmsg** because it provides:

- an extensible mechanism for encapsualting rendezvous and IP traffic trace (flow) data
- a means of transmitting streams to distributed encapsulation and online analysis elements
- a serialized file format for offline analyses
- a scripting interface to build prototype components and perform ad hoc analyses

Rendezvous-annotated Flow Export



Rendezvous-annotated Flow Export (1)

[2011-06-08 21:52:26.00000000] [7:1 WISC nfdump]

- ts: 1307569945
- te: 1307569945
- td: 0.064000
- sa: 203.0.113.71
- da: 192.0.2.32
- sp: 80
- dp: 55983
- pr: 6
- ibyt: 396630

```
snamed: CLIENT_DNS_NAMED
```

```
sn: static.ak.facebook.com
```

```
ip_version: IPV4
```

Rendezvous-annotated Flow Export (2)

[2011-06-08 20:14:11.000000000] [7:1 WISC nfdump]

- ts: 1307564050
- te: 1307564050
- td: 0.064000
- sa: 2001:0db8::face:b00c:0:3
- da: 2001:0db8::2:1
- sp: 443
- dp: 53646
- pr: 6
- ibyt: 34297
- snamed: INFERRED_DNS_NAMED
- sn: *.facebook.com.
- sn_sample: de-de.facebook.com.
- sn_sample: check6.facebook.com.
- sn_sample: ar-ar.facebook.com.
- ip_version: IPV6

Rendezvous-annotated Flow Export (3)

[219] [2011-06-08 00:11:10.00000000] [7:1 WISC nfdump]

- ts: 1307491869
- te: 1307491869
- td: 0.128000
- sa: 2001:0db8::2:1
- da: 2001:0db8:fff4::79
- sp: 56451
- dp: 80
- pr: 6
- ibyt: 849

```
dnamed: INFERRED DNS NAMED
```

```
dn: *.
```

```
dn_sample: rss.slashdot.org.
```

```
dn_sample: www.beantownbloggery.com.
```

ip_version: IPV6

World IPv6 Day Performance Study: Trace Data Characteristics

Characteristic	Count
Trace duration	24 hours
DNS query responses	~14.2M
DNS IPv4 client addresses	2028
DNS IPv6 client addresses	23
DNS AAAA queries	~114.3K
DNS AAAA NOERROR responses	~6.2K
Flows - IPv4	~58.8M
Flows - IPv6	~2.4M

World IPv6 Day: Popular IPv6 FQDNs

Rank	FQDN
1	www.google.com.
2	www.google-analytics.com.
3	www.facebook.com.
4	ssl.gstatic.com.
5	safebrowsing.clients.google.com.
6	mail.google.com.
7	safebrowsing-cache.google.com.
8	clients1.google.com.
9	www.youtube.com.
10	view.atdmt.com.
11	ajax.googleapis.com.
12	news.google.com.
13	maps.google.com.
14	ssl.google-analytics.com.
15	addons.mozilla.org.
16	docs.google.com.
17	chatenabled.mail.google.com.
18	translate.google.com.
19	mail-attachment.googleusercontent.com.
20	sites.google.com.

Results

- Service domain names, e.g.:
 - Facebook: 950 FQDNs with suffix "facebook.com" ~618K IPv4 flows, ~128K IPv6 flows

Gmail: "gmail.com", "mail.google.com",
 "www.gmail.google.com"
 ~785K IPv4 flows, ~463K IPv6 flows

- IPv4 and IPv6 service asymmetries, e.g.:
 - Facebook: "*.channel.facebook.com" (chat) queries were resolved as A but not AAAA
 - Gmail: "imap.gmail.com", "smtp.gmail.com" queries were resolved as A but not AAAA
- Active host counts and service flow rates ...

Facebook Active Client IP Addresses



Gmail Active Client IP Addresses



Facebook WWW Flow Bit Rates (detail)



Gmail WWW Flow Bit Rates (detail)



Facebook WWW Flow Bit Rates



Gmail WWW Flow Bit Rates



Summary

• Contribution:

A passive method and framework to examine services' performance on IPv6 and IPv4, accommodating dual-stack/hybrid client access and exposing services' IP protocol version configurations and asymmetries

 Nascent related work by others: "Inferring Internet Server IPv4 and IPv6 Address Relationships" [Beverly, Berger et al. 2013]

Assessing Performance of Internet Services on IPv6

FIN



Residential: Domain Popularity



Aggregate Traffic: named & unnamed



Aggregate Traffic by Domain Name

