A Day's Worth of Changes to Fully Qualified Domains On The Global Internet

Abstract. This study analyzes a day's worth of changes to fully qualified domain names (n= 29,620,807), collected from over 450 sensors worldwide. The rate of new domain creation is estimated. The distribution of resource record types is found. We also identified multiple types of high frequency FQDN (base domain, record type)-tuples, and highlighted easily identified security-relevant domain names.

1 Introduction

The Farsight Security, Inc., Security Information Exchange (FSI SIE) offers researchers and commercial customers access to DNS traffic donated by over 450 sensor-node operators. This real-time data allows those with access to conduct cross-sectional and longitudinal studies of global Internet DNS traffic.

New domains, and newly-changed domains, are known to be particularly salient. Responding to market interest in new domain names, FSI has offered a production feed of newly-seen Internet-wide base domain names, but has not offered a feed of newly-seen Internet-wide host names (or "fully qualified domain names" (FQDNs)). FSI has also not previously offered a production feed that traces changes to existing DNS data.

This paper characterizes data gleaned from a new experimental FSI SIE channel, looking at a day's worth of newly-seen FQDNs on the global Internet, including a day's worth of changes to FQDNs.

2 Data Collection

Farsight staff provided newly-observed FQDN/newly-changed FQDN data for a 24 hour period running from 0600 GMT on 2015/05/21 through 0599 GMT on 2015/05/22. A total of 29,620,807 observations were received, distributed over time as shown in Figure 1.

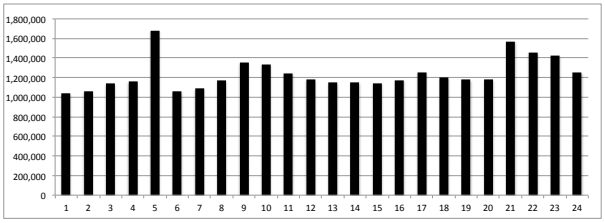

Fig. 1. Observations per hour during 24 hour study period

That volume of traffic (in observations (or "payloads") per second) is consistent with routine RRDtool graphs provided for another day, running on average roughly 360 payloads per second ((362.42 payloads per second)*(60*60*24 seconds per day)=31,313,088 payloads per day.

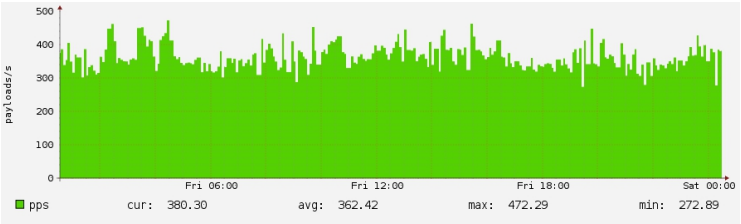

Fig. 2. Payloads per second traffic for an alternative representative day

3 Data Format

Observations were distributed at SIE in a new NMSG format known as "SIE newdomain." A couple of sample observations from that data (represented here in "presentation" or "display" format) look like:

Sample Observation Number 1:

```
domain: ncpqa.cn.
time_seen: 2015-05-21 07:56:35
bailiwick: ncpqa.cn.
rrname: qqgfx.ncpqa.cn.
rrclass: IN (1)
rrtype: A (1)
rdata: 104.151.50.125
new_domain: False
new_rrname: True
new_rrtype: True
new_rr: True
new_rrset: True
```

This record is an example of a new fully qualified domain name ("qqgfx.ncpqa.cn") that's part of a base domain FSI had previously seen ("ncpqa.cn"). The DNS record that was observed by an FSI sensor node was an "A" record mapping qqgfx.ncpqa.cn to the IP address (104.151.50.125). Because this is a new rrname, the resource record type (rrtype), the resource record itself (rr), and the resource record set (rrset) will also be new for that rrname.

Sample Observation Number 2:

```
domain: cdn13.com.
time_seen: 2015-05-21 07:57:40
bailiwick: cdn13.com.
rrname: a9-19.clients.cdn13.com.
rrclass: IN (1)
rrtype: A (1)
rdata: 206.54.168.3
rdata: 206.54.168.7
rdata: 206.54.168.11
rdata: 206.54.168.15
rdata: 206.54.168.16
rdata: 206.54.168.38
rdata: 206.54.168.39
rdata: 206.54.168.41
rdata: 206.54.168.44
rdata: 206.54.168.46
new_domain: False
new_rrname: False
new_rrtype: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rr: False
new_rrset: True
```

Like observation number 1, observation number 2 is another "A" record. Unlike observation number 1, this rrname was set up by its owner to resolve to multiple IP addresses. While the rrtype and each individual resource records aren't new for this previously seen rrname, the resource record set as a whole is different from previously observed resource record sets for this rrname.

4 New Domains Seen In The Day's Worth of Data

Out of 29,620,807 total observations, 176,366 (0.595%) represented identification of new base domains.

Assuming a uniform rate of new domain generation, this would imply an annual rate of new domain creation of (176,366 domains per day*365.25 days per year)=64,417,681 domains per year. That rate is over three times higher than the net growth in new domains reported by VeriSign. [Verisign, 2015] The magnitude of that discrepancy is large enough to merit further definitive study, but some potential reasons for that discrepancy may be a combination of:

– The fact that the Farsight data focused solely on new domains (without adjusting for expiring or deleted domains), while VeriSign is believed to be looking at net growth (domains added less domains expired or deleted).
– Farsight counts both newly registered domain names plus new domains created under effective top level domains as determined by the Public Suffix List [Mozilla, 2015] while VeriSign does not consider new domains created under Public Suffix List effective top level domains.
– There may be genuine changes in the rate of domain name creation, potentially associated with the ICANN new gTLD program, increased uptake of IDNs (internationalized domain names), etc.
– While FSI has been tracking new domains since June 2010, FSI is continually adding new sensor nodes. Some of the newly-detected new domains may actually have previously existed, but may be just now seen by Farsight as their coverage continues to improve.

5 Newly Observed Fully Qualified Domain Names

7,778,302 observations (26.26%) represented identification of new rrnames (e.g., new FQDNs). Extrapolating, this translates to an annualized rate of new FQDN creation of 7,778,302*365.25=2.84 billion new FQDNs per year.

6 Observed Resource Record Types

While "A" records (which map domain names to IPv4 addresses) are the most well-known type of DNS record – and the most common type of DNS record seen in our day's worth of data – "A" records were not the only record type seen:

Table 1: Observations broken down by resource record type

Observations	% of Obs	Record Type & Code
16,964,386	57.27%	A (1)
9,460,957	31.94%	SOA (6)
1,745,213	5.89%	CNAME (5)
714,677	2.41%	NS (2)
259,468	0.88%	PTR (12)
204,785	0.69%	MX (15)
149,771	0.51%	TXT (16)
100,424	0.34%	AAAA (28)
18,140	0.06%	NULL (10)
2,393	0.01%	SRV (33)
440	<0.01%	SPF (99)
77	<0.01%	WKS (11)
59	<0.01%	<UNKNOWN>(1169)
7	<0.01%	DNAME (39)
4	<0.01%	LOC (29)
3	<0.01%	HINFO (13)
1	<0.01%	<UNKNOWN>(4652)
1	<0.01%	<UNKNOWN>(4097)
1	<0.01%	RP (17)
29,620,807	**100.00%**	

7 High Frequency (base domain, type)-tuples

Succinctly yet comprehensively characterizing over 29 million observations poses distinct practical challenges, and any attempt will necessarily be incomplete. However, there are some obvious points that particularly merit comment. We discuss those in Sections 8-11.

8 Content Distribution Networks With Interleaved Location-Dependent DNS Answers

Many DNS names resolve consistently regardless of who's resolving them or where those questions may be originating. Some authoritative name servers, however, may NOT return the same results for all query sources.

For instance, in an effort to minimize query latency, some content distribution networks (CDNs) may intentionally provide different answers for a query depending on the query origin. A query from a North American user may be sent to a nearby North American server for action, while another query for the same domain name at the same moment in time (but from an Australian user) may be sent to a server in Australia, instead.

Farsight sees those interleaved queries and responses because they have many different sensors located all around the world. Those interleaved queries may appear to be "changes" even if the response that any particular client see in utterly invariant over time, simply because the CDNs answer to a given query will vary depending on the source of that query, or query origin plus other factors (such as load balancing considerations).

The names with the highest number of changes are, in fact, consistently associated with CDNs. The (rrname, rrtype) tuples that had the highest frequencies (over 300,000 records for the period of observation) were virtually all CDN-related. See table 2.

Table 2: RRnames seeing the largest number of daily "changes:" unique individual rrnames with N(obs) >300,000

Observations	RR Name	Type & Code
564,491	stun.client.akadns.net.	A (1)
521,283	dr-asia.skype-cr.akadns.net.	A (1)
464,072	dr.skype-cr.akadns.net.	A (1)
329,899	dsn4.skype-dsn.akadns.net.	A (1)
329,623	dsn12.skype-dsn.akadns.net.	A (1)
329,514	dsn6.skype-dsn.akadns.net.	A (1)
329,481	dsn15.skype-dsn.akadns.net.	A (1)
329,213	dsn10.skype-dsn.akadns.net.	A (1)
329,128	dsn3.skype-dsn.akadns.net.	A (1)
329,060	dsn1.skype-dsn.akadns.net.	A (1)
328,884	dsn2.skype-dsn.akadns.net.	A (1)
328,784	dsn14.skype-dsn.akadns.net.	A (1)
328,709	dsn9.skype-dsn.akadns.net.	A (1)
328,637	dsn0.skype-dsn.akadns.net.	A (1)
328,600	dsn13.skype-dsn.akadns.net.	A (1)
328,542	dsn8.skype-dsn.akadns.net.	A (1)
328,502	dsn5.skype-dsn.akadns.net.	A (1)
328,495	dsn7.skype-dsn.akadns.net.	A (1)
328,238	dsn11.skype-dsn.akadns.net.	A (1)
321,825	px-lax007.quantserve.com.akadns.net.	A (1)

9 Frequently Updated Zones/High Frequency SOA Records

Another class of "frequently changing" observations were associated with SOA (Start of Authority) records.

SOA records are used in the DNS to specify the maintainer of a DNS zone, the zone's TTL (time-to-live) values for DNS caching-related purposes, and a serial number identifying the current version of the zone file. That serial number is normally incremented whenever the zone file is updated. That change in serial number is sufficient to trigger a change detection in the current dataset. Thus, it is not surprising that a second category of high frequency (rrname, type)-tuples is associated with SOA records for frequently updated zones. If a zone were to be updated every second, that would imply (60 seconds/minute)*(60 minutes/hour)*(24 hours/day)=86,400 changes (one for each second of the day). The SOA values we observed approximate those values. See table 3.

Table 3. RRnames seeing the largest number of daily "changes:" Top 15 unique individual (rrname, type=SOA) observations

Observations	RR Name	Type & Code
83,093	akadns.net.	SOA (6)
82,739	g.akamai.net.	SOA (6)
82,716	g.akamaiedge.net.	SOA (6)
82,690	da1.akamai.net.	SOA (6)
82,688	w28.akamai.net.	SOA (6)
82,649	w29.akamai.net.	SOA (6)
82,635	b.akamai.net.	SOA (6)
82,622	w22.akamai.net.	SOA (6)
82,610	b.akamaiedge.net.	SOA (6)
82,608	g2.akamai.net.	SOA (6)
82,579	d.akamai.net.	SOA (6)
82,550	w27.akamai.net.	SOA (6)
82,546	g1.akamai.net.	SOA (6)
82,543	a.akamaiedge.net.	SOA (6)
82,528	w23.akamai.net.	SOA (6)

10 Base Domains With A Large Number of Unique rrnames

Many domains will naturally have only a single rrname show up in this DNS changes data. For example, a domain name might have traditionally been on an IP obtained from one provider, only to change to a new IP when the domain begins to be hosted by a new provider. Such a domain will be characterized by being seen for a long time on one IP, and then persistently on a new IP, a

change that will be reflected in the rrname having a single DNS change. Other observations consist of base domains that have large numbers of unique rrnames, where those new rrnames are only seen once. These use-it-once-and-discard it unque "disposable rrnames" may be a sign that DNS is being used as:

- A tracking mechanism (e.g., for per-web-page or per-email-message tracking links),
- A data exfiltration mechanism (e.g., for data-over-DNS surreptitious data transfers),
- An anti-monitoring mechanism (e.g., to help keep any single FQDN appearing to be too "hot" or "active"), or in
- Some other unconventional manner.

If we collapse observed rrnames using the Effective TLD/Public Suffix List, we are left with a list of most-frequently-observed base domains. We'll exclude Akamai-related domains (those have already been prominently featured in tables 2 and 3, above), plus Amazon-related domains (another obviously-massive provider-at-scale), as well as some CDN-related domain names (other than Akamai) that might otherwise also have been included (e.g., cdn13.com, fbcdn.net, and cdn77.net). We'll also exclude a domain associated with a site that's operating a sensor node for FSI SIE to avoid disclosing that data source, consistent with FSI's terms of service/non-disclosure agreement requirements.

Table 4. Selected Base Domains With Relatively Large Numbers of Observations

Obs	Base Domain	Whois	Registrar
987,688	yahoodns.net	regular	MarkMonitor
544,784	tumblr.com	regular	MarkMonitor
373,145	telemetryverification.net	regular	Tucows
263,528	rssing.com	regular	Key Systems
260,600	parse.com	regular	MarkMonitor
234,518	mgm86800.com	private	Godaddy
210,266	mgm001.com	regular	Godaddy
204,723	mgm86877.com	private	Godaddy
199,374	mgm002.com	regular	Godaddy
158,622	spotilocal.com	regular	DomaininfoAB
130,961	surfeasy.mobi	regular	Godaddy
94,776	adnxs.net	regular	MarkMonitor
74,757	vkrugudruzei.ru	web-based	RU-Center-RU
66,913	ns1p.net	private	Name.com
63,571	sekindo.com	regular	Dyn.com
62,474	optinre.ru	private person	Salenames-RU
61,335	1drv.com	regular	MarkMonitor
55,232	nessus.org	regular	Network Sol.
54,505	spampoison.com	regular	Enom
53,625	incapsecuredns.net	regular	Godaddy
50,841	seagateshare.com	regular	MarkMonitor
48,574	greatrelating.com	private	Melbourne IT
47,683	worldssl.net	regular	Enom
47,455	mailguard.com.au	web-based	Melbourne IT
46,723	spotify.com	regular	DomaininfoAB
45,553	wd2go.com	regular	CSC
41,207	geoadnxs.com	regular	MarkMonitor
40,368	dyndns.org	regular	Tucows
37,704	imdb.com	regular	MarkMonitor
36,034	emltrk.com	regular	Enom
35,750	mgm86855.com	private	Godaddy
33,913	bugun.in	no street address	Name.Com
32,557	websamsung.net	regular	WhoisNetworks
31,232	audible.com	regular	CSC
30,757	notifygate69.ru	private person	R01-RU
30,026	notifygate72.ru	private person	R01-RU
29,871	notifygate70.ru	private person	R01-RU
29,782	notifygate71.ru	private person	R01-RU
29,346	notifygate73.ru	private person	R01-RU

If we look at actual rrnames associated with one of those domains, such as notifygate69.ru, we see a pattern consistent with programmatic domain name generation:

```
abbadided.notifygate69.ru.
aberdeenn.notifygate69.ru.
accordanceg.notifygate69.ru.
accrescei.notifygate69.ru.
accruementg.notifygate69.ru.
addictivep.notifygate69.ru.
adjudgeri.notifygate69.ru.
adrenalonea.notifygate69.ru.
adventuredl.notifygate69.ru.
affirmablyy.notifygate69.ru.
afterpotentialt.notifygate69.ru.
aggravatives.notifygate69.ru.
agricolitec.notifygate69.ru.
airohydrogeny.notifygate69.ru.
aistopodesm.notifygate69.ru.
aluminasy.notifygate69.ru.
amalfianf.notifygate69.ru.
amaryllideousm.notifygate69.ru.
ambidextrousv.notifygate69.ru.
ambritev.notifygate69.ru.
amentiaq.notifygate69.ru.
amicablenessb.notifygate69.ru.
aminoaceticz.notifygate69.ru.
amortizingx.notifygate69.ru.
amphicondylac.notifygate69.ru.
amuttera.notifygate69.ru.
amygdalotomyt.notifygate69.ru.
amyloidala.notifygate69.ru.
anabiosisu.notifygate69.ru.
anaglyptographyt.notifygate69.ru.
anaphalish.notifygate69.ru.
anchitherioidr.notifygate69.ru.
annexationistv.notifygate69.ru.
anoegenetico.notifygate69.ru.
antanaclasisk.notifygate69.ru.
anthracnosep.notifygate69.ru.
anthropoidse.notifygate69.ru.
antiberiberina.notifygate69.ru.
antichlora.notifygate69.ru.
antidactylj.notifygate69.ru.
antidiastasep.notifygate69.ru.
antirentismp.notifygate69.ru.
antitaxz.notifygate69.ru.
anywhenf.notifygate69.ru.
[etc]
```

For context, at the time this paper was written, the domain notifygate69.ru is currently listed in the Spamhaus Domain Blocklist (DBL); see http://www.spamhaus.org/query/dbl?domain=notifygate69.ru and on SURBL.

11 Domain Names of Special Security Relevance

Access to a dataset with details about DNS changes also can be useful for identifying new special domain names, such as domains that may be associated with brand infringement or phishing.

For example, while "paypal.com" is the well known brand of a leading online payments company, a number of other rrnames also incorporate that brand name. Looking for the string "paypal" in rrnames seen in the DNS changes data, we saw 274 unique domain names incorporating that term, besides paypal.com itself. We wouldn't want anyone to accidentally visit these dubious sites, so we've reversed a particularly dubious-looking sampling of those rrnames and replaced the dots in those names with [dot] for additional protection against accidental visits:

```
bg [dot] paypal-topup
br [dot] com [dot] construcasarj [dot]
     notice-of-changes-to-the-paypal-user-agreement
co [dot] iinc [dot] paypal [dot] limited [dot] secure [dot] login
com [dot] 4paypal
com [dot] actiumweb [dot] solmallorca2 [dot] paypal-account-confirmation
com [dot] activated-paypal-cash
com [dot] appsmyway [dot] paypall [dot] account [dot] update
com [dot] confirmpaypalls
com [dot] confirm-your-account [dot] verification [dot] paypal
com [dot] contactservicepaypal
com [dot] customer-paypal-update
com [dot] engineerslabltd [dot] informations [dot] verification [dot]
      paypal-community [dot] www
com [dot] mashangqifei [dot] update-your-information-paypal-account
com [dot] myskmg [dot] update-paypal-account
com [dot] myzazzlestore [dot] paypal-update-your-info-2015-2016-security
com [dot] new-lineapparel [dot]
      paypal-service-support-update-your-account-information-security [dot] www
com [dot] paypalaccountupdating
com [dot] paypal-com-cgi-bin-webscr-login-access
com [dot] paypal-new-security
com [dot] special-batu-akik [dot]
      com-support-id-verifyy-accountt-limitation [dot] m-paypal
com [dot] verify-paypal-center
org [dot] credit-card-processing-services [dot] confirmation-account-paypal
us [dot] validatepaypal [dot] www
```

12 Conclusion

We've described and characterized a unique data set about global DNS changes taking place at the FQDN-level based on a day-long sample of over 29 million observation.

We've identified a number of scenarios that present special challenges for any effort devoted to tracking such changes, including intentionally inconsistent "A" records used by content distribution networks, highly volatile DNS zones, and programmatically generated domain names.

We've also provided a concrete example of how FQDN-level change data can be used to identify probable phishing domains for review and appropriate action.

Acknowledgments

This analysis would not have been possible without Security Information Exchange data provided by Farsight Security, Inc.

References

1. "The Domain Name Industry Brief," Vol. 11, Issue 4, January 2015, https://www.verisigninc.com/assets/domain-name-report-january2015.pdf
2. "The Public Suffix List," https://publicsuffix.org/