

# Multivariate Solutions to Emerging Passive DNS Challenges





# Agenda

- Introduction
  - Passive DNS, Including Times When Passive DNS May Not Work Well
- Overcoming Obfuscation
  - Pillz Spam Example
  - Brand Protection/Knock Off Jerseys Example Scheduled Controlled Substances
  - Working A Kelihos Botnet-Related Spam Example
- Multivariate Solutions
- Conclusion



#### I. Introduction:

Passive DNS, Including Times When Passive DNS May Not Work Well



### **How Passive DNS Normally Works...**

#### [From the POV of a security analyst]

- Start with a known/observed "bad data point"
  - Domain name
  - Nameserver
  - IP address/CIDR
  - ASN (→ CIDRs)
- Use Passive DNS to find other IPs or domain names that share the same resources as our evil clue
- Leverage reputation locality ("guilt by association"), but carefully review what you've found



### **UNIvariate Approaches**

 Use a single point of commonality as a way to identify related domains...

- Same exact IP?
- Same exact nameserver?
- Same exact domain name used over time (if you're interested in the set of IPs that a name's been using)
- Each relies on a **single** attribute, **exactly matched**.



### Simple pDNS Works GREAT When...

- Lots of related domains coexist on a single IP (or small CIDR block), with no innocent 3<sup>rd</sup> party domains
- Many related domains use the same set of dedicated name servers, with no innocent 3<sup>rd</sup> party domains
- The bad guy is apparently stubbornly fond of a favorite domain, despite being kicked off provider after provider after provider



### Times When Simple pDNS <u>Doesn't</u> Work

- **ZERO** interrelated data points e.g., "lone wolf" domain names, IP addresses, name servers, etc.
- TOO many related resources
- Related bad guy resources are <u>comingled</u> <u>inextricably</u> with innocent 3<sup>rd</sup> party resources.
- Bad Guy "Hit and run" scenarios



#### **Lone Wolf Scenario**

#### The cybercriminal reuses NOTHING across sites

- Every IP address used to send spam or host content is totally unrelated to any other IPs the criminal uses
- Every domain name is registered using:
  - A diverse assortment of registrars, one or two at a time
  - Using unique name servers (installed and operating on unique IPs)
  - Unique/fictitious (or concealed) POC details
  - Unique (or anonymous) payment details

#### Each site uses:

- different brand names
- different images
- different written text
- different payment processors, etc.



### **Large Heavily-Multiplexed Resources**

- **Example #1:** A registrar's shared nameservers support 100,000 customer domains, good, bad and ugly.
- Example #2: Bad domains hide behind a shared reverse-proxy service (i.e. CloudFlare) -- thereby concealing a domain's actual IP addresses



#### **Poorly Documented Resource Assignments**

- Example #1: Provider fails to document IP reassignments/reallocations in IP Whois or rWhois, and an abuser repeatedly moves (or is moved) around a single large network block, or among multiple smaller blocks.
- Example #2: Whois POC details are concealed by a Whois proxy/privacy service



# **II. Overcoming Obfuscation**



### Work Around It, Or Strip It Entirely

 Look for other characteristics that may not be obfuscated, or seek to strip away anonymity.

#### For example:

- If nameservers service a large number of domains, and thus are not a useful attribute to try to follow, look at the IP address(es) the bad domain is hosted on, instead.
- If a domain is demonstrably engaged in phishing or other clearly illegal behavior, some privacy/proxy protection services have terms of service which allow the provider to unilaterally strip privacy protections.

### F < R S I G H T

#### **Strategies For Overcoming Reverse Proxies**

- With Reverse Proxies, everything seems to "live on the reverse proxy's IP addresses"
- Carefully scrutinize non-A/non-AAAA DNS records that may be present (e.g., MX, TXT, etc.)
- Reverse proxy operators are also potentially a <u>terrific</u> target by law enforcement



### **Bad Guys Deobfuscate Good Guys, Too**

- "Performance Marketing" URLs are encoded URLs, unique to each specific recipient
- Because each URL is unique to each recipient, visiting the URL (typically to investigate the site being spamvertised) means:
  - Confirming you've opened the message and clicked through (establishing a potential argument that you've "opted-in")
  - May result in you "using-up" a URL coded for one-time-use (try the same URL a 2nd or 3<sup>rd</sup> time? It may go nowhere)
  - Forwarding "sanitized" spamples in complaints may yield URLs that simply don't work, or which work "misleadingly."
  - Forwarding "raw spamples in complaints "outs" your spam collection infrastructure and may result in "list washing."



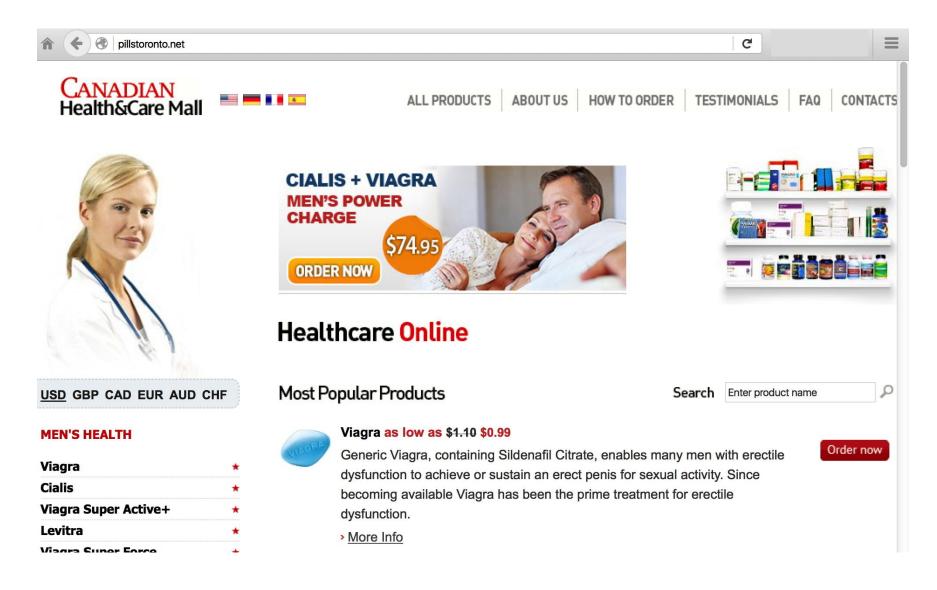
### **II-a. Overcoming Obfuscation:**

**Pillz Spam Example** 

Demonstrates Use of Historical Passive DNS Data to Overcome Reverse Proxy Usage



### **An Anti-Spam Example: Pillz**



### **Using Pre-"Reverse-Proxy-fication" Data**

```
$ dnsdb query.py -r pillstoronto.net/a
;; bailiwick: pillstoronto.net.
       count: 548
;; first seen: 2015-06-07 12:57:11 -0000
;; last seen: 2016-01-19 00:46:36 -0000
pillstoronto.net. IN A 104.24.126.91 ← Cloudflare now
pillstoronto.net. IN A 104.24.127.91 ← Cloudflare now
[BUT, EARLIER, WE'D SEEN...]
;; bailiwick: pillstoronto.net.
;; count: 5,568
;; first seen: 2012-09-03 19:53:45 -0000
;; last seen: 2013-09-11 19:41:57 -0000
pillstoronto.net. IN A 188.72.228.107 ← NOT Cloudflare
;; bailiwick: pillstoronto.net.
       count: 4,965
;; first seen: 2013-09-11 21:22:24 -0000
;; last seen: 2015-06-07 09:08:03 -0000
pillstoronto.net. IN A 80.67.3.104 ← NOT Cloudflare
```



#### The Guys Behind These Guys Go Way Back

"EvaPharmacy (previously known as Bulker.biz) is the organization which sponsors spammers to promote sites within what has previously been referred to as the Yambo Financials group of web properties. These include My Canadian Pharmacy, International Legal RX, Canadian Health&Care Mall, US Drugs, Canadian Family Pharmacy, Canadian Family Pharmacy, Toronto Drug Store, RxExpressOnline, RxMedications and others. This was learned from postings on bulkerforum.biz by username "ebulker", who would invite users to promote for their properties. [...] Eva Pharmacy brand websites were first discovered in **2007** loading content from Bulker.biz sites."

http://fraud-reports.wikia.com/wiki/EvaPharmacy [emphasis added]



### II-b. Overcoming Obfuscation:

### **Brand Protection/Knock Off Jerseys Example**

# Illustrate Use of MX Record Info To Overcome Reverse Proxy Usage



### **Context for This Example**



www.ice.gov/news/releases/operation-team-player-nets-more-37-million-fake-merchandise

#### INTELLECTUAL PROPERTY RIGHTS

02/13/2014

#### 'Operation Team Player' nets more than \$37 million in fake merchandise

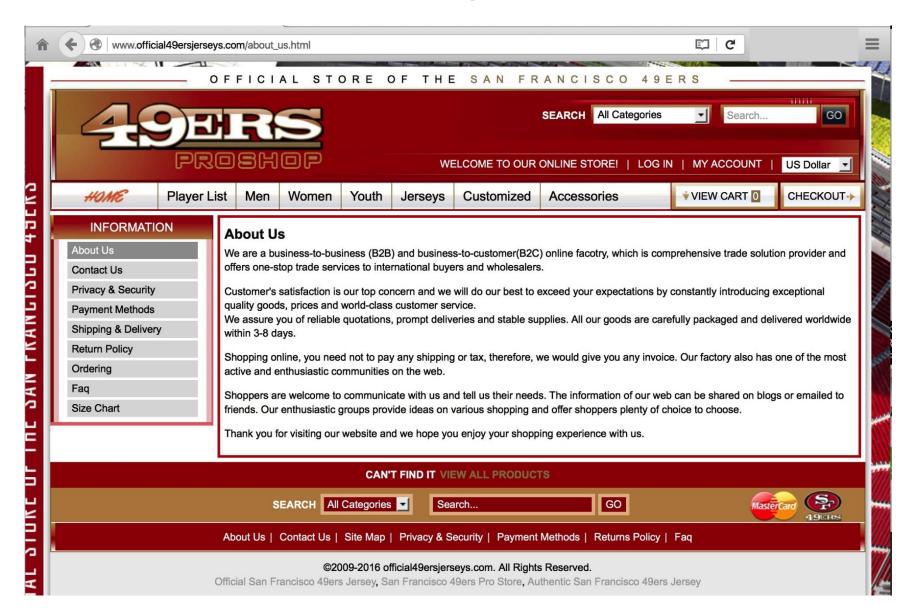
More than 70 people arrested; over 5,000 websites seized in coordination with NFL

WASHINGTON — Federal officials announced Thursday the final record-breaking results of Operation Team Player, the nationwide law enforcement effort aimed at combatting counterfeit sports merchandise.

Special agents from U.S. Immigration and Customs Enforcement's (ICE) Homeland Security Investigations (HSI) teamed with officers from U.S. Customs and Border Protection (CBP) to target, seize and investigate criminal businesses smuggling international shipments of counterfeit merchandise as it entered the United States. Agents also targeted warehouses, stores, flea markets, online stores and street vendors. The operation, which began in June, netted 397,140 items including fake jerseys, ball caps, T-shirts, jackets and other souvenirs. The items had a manufacturer's suggested retail price (MSRP) of more than \$37.8 million – more than the previous six Super Bowl enforcement efforts combined.



### Is This Really The "Official Store?"





#### **Compare Two Domain Whois Entries**

Domain Name: official49ersjerseys.com

[...]

Create Date: 2015-09-03 14:24:36

[...]

**Registrar: SHANGHAI MEICHENG** 

**TECHNOLOGY INFORMATION** 

**DEVELOPMENT CO., LTD.** 

[...]

Registrant Name: shao nian

Registrant Organization: shao nian

Registrant Street: Shang Hai Shi Qu

Registrant City: shanghaishi

Registrant State/Province: shanghai

Registrant Postal Code: 123123

Registrant Country: CN

Registrant Phone: +86.021**1231231** 

Registrant Fax: +86.0211231231

Registrant Email: cj2015tit@126.com

[etc]

Domain Name: nflshop.com

[...]

Updated Date: 2015-07-14T04:00:24-0700

**Creation Date: 1999**-02-01T00:00:00-0800

Registrar: MarkMonitor, Inc.

[...]

Registrant Name: NFL Enterprises LLC

Registrant Organization: NFL Enterprises LLC

Registrant Street: 345 Park Ave.,

Registrant City: new york

Registrant State/Province: ny

Registrant Postal Code: 10017

Registrant Country: US

Registrant Phone: +1.2124502000

[...]

Registrant Email: dns admin@nfl.com

[etc]

Which of these two domains do YOU think is the real official NFL jersey shop?



### Following MX Records as DNS Clues

```
$ dig official49ersjerseys.com +short
104.27.142.198  Hidden behind Cloudflare
$ dig official49ersjerseys.com mx +short
0 dc-96d9f219.official49ersjerseys.com.
$ dig dc-96d9f219.official49ersjerseys.com +short
Do the "regular Passive DNS dance" from that point...
$ dnsdb query -i 107.155.198.200 -p json | jq -r .rrname |
2nd-level-dom | sort -u
cheapcustomjerseysonline.com.
dallascowboymall.com.
dallascowboysmalls.com.
[etc]
dnsdb query (c lang)? see https://github.com/dnsdb/dnsdb c
Get jq from https://stedolan.github.io/jq/
```

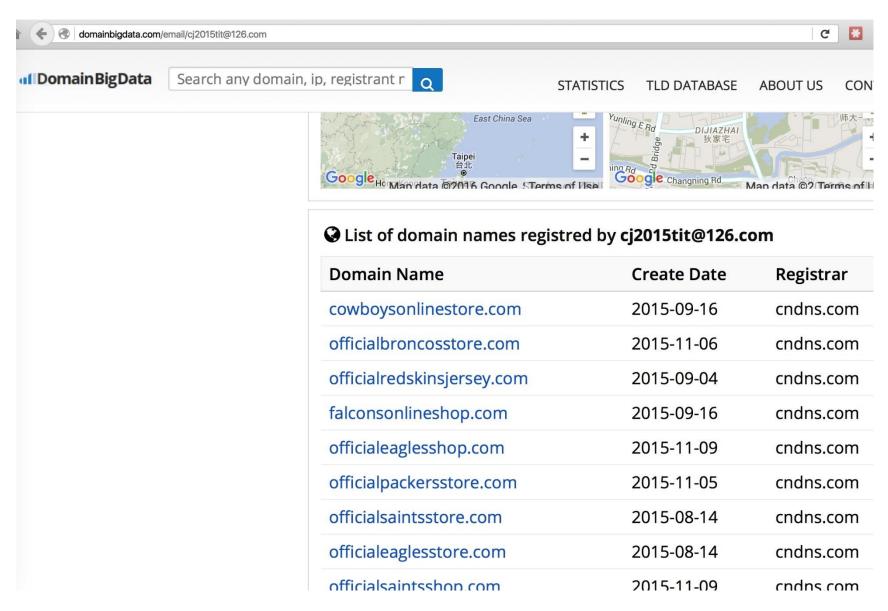


#### [Aside: "2nd-level-dom" is Just a Small Perl Script]

```
#!/usr/bin/perl
use strict;
use warnings;
use IO::Socket::SSL::PublicSuffix;
my $pslfile = '/usr/local/etc/effective tld names.dat';
my $ps = IO::Socket::SSL::PublicSuffix->from file($pslfile);
my $line;
foreach $line (<>) {
        chomp ($line);
        my $root domain = $ps->public suffix($line,1);
        printf( "%s.\n", $root domain );
Get effective tld names.dat from
https://publicsuffix.org/list/effective tld names.dat
```



#### Got an Email? You Can Follow That, Too





### II-c. Overcoming Obfuscation:

**Scheduled Controlled Substances** 

Illustrates Use of TXT Record Info To Overcome Reverse Proxy Usage



#### **Anabolic Steroids**

www.buysteroidonline.com Checkout **Payment** Shipping Contact Us Home **Products** PRODUCT CATEGORIES Online shop, where you can buy all kinds of anabolic steroids. You can easily buy all kinds of hormones; anabolic Injectable steroids like Deca Durabolin, Sustanon, Winstrol, Equipoise; Human Growth Hormone, and androgenic hormone Vitagon or Pregnyl and other hormonal drugs. You can buy oral steroids Anabol, Oxandrolone and Trenbolone from this authentic shop. The Buysteroidonline sell quality and genuine steroid and it collects the genuine product from the manufacturing company. You can easily collect original steroids with confidence. We have all genuine and latest anabolic steroids present in the market. You can easily buy Anti-Estrogen, Anti Depressants, Sexual Enhancement, Weight Loss and Skin & Hair etc. We give all information like effect, dosages and side effect. You can find any desire drug from our online shop. Please ask us any question and feel free to contact us any time. The payment process of our **BEST SELLERS** shop is easiest. You can pay the price of product with different currencies and you can choose the currency from the top right site of our home page. If your order over \$500, you can get free shipping worldwide.



#### **Anabolic Steroids Are Schedule III**

SUBSTANCE	DEA NUMBER	CSA SCH	NARC	OTHER NAMES
Alpha-methylfentanyl	9814	I	Υ	China White, fentanyl
Alpha-methylthiofentanyl	9832	- 1	Y	China White, fentanyl
Alpha-methyltryptamine	7432	ı	N	AMT (Positional Isomer: N-Methyltryptamine)
Alphaprodine	9010	II	Υ	Nisentil
alpha-pyrrolidinobutiophenone (α-PBP)	7546	1	N	1-phenyl-2-(pyrrolidin-1-yl)butan-1-one)
alpha-pyrrolidinopentiophenone (α-PVP)	7545	I	N	α-pyrrolidinovalerophenone, 1-phenyl-2- (pyrrolidin-1-yl)pentan-1-one)(Positional isomers: 4-methyl-α-pyrrolidinobutiophenone (4-MePBP), 1-phenyl-2-(piperidin-1-yl)butan-1-one)
Alprazolam	2882	IV	N	Xanax
AM-2201 (1-(5-Fluoropentyl)-3-(1-naphthoyl) indole)	7201	1	N	AM-2201
AM-694 (1-(5-Fluoropentyl)-3-(2-iodobenzoyl) indole)	7694	1	N	AM-694
Aminorex	1585	1	N	has been sold as methamphetamine
Amobarbital	2125	II	N	Amytal, Tuinal
Amobarbital & noncontrolled active ingred.	2126	III	N	
Amobarbital suppository dosage form	2126	III	N	
Amphetamine	1100	П	N	Dexedrine, Adderall, Obetrol
Anabolic steroids	4000	III	N	"Body Building" drugs



#### **Schedule III Carries Stiff Penalties**

**Trafficking** (Unlawful distribution, possession with intent to distribute, manufacture, importation and exportation, etc. (21 U.S.C. 841, 960, 962, and 46 U.S.C. 70506), Any Weight

1<sup>st</sup> Offense: \$500,000/\$2.5 million Up To 15 years

2<sup>nd</sup> Offense: \$1 million/\$5 million Up to 30 years

Fines shown are for an individual/for defendants other than an individual. Terms are maximum periods of incarceration.

Many other related offenses and penalties are summarized in "Drug Offenses: Maximum Fines and Terms of Imprisonment for Violation of the Federal Controlled Substances Act and Related Laws", https://www.fas.org/sgp/crs/misc/RL30722.pdf

#### Following A TXT Record as A DNS "Clue"

```
$ dig buysteroidonline.com +short
104.28.0.126  Hidden behind Cloudflare
104.28.1.126  Hidden behind Cloudflare
$ dig buysteroidonline.com txt +short
"v=spf1 +a +mx +ip4:193.111.62.68 ~all"
$ dig buysteroidonline.com mx +short
0 dc-ce20a397.buysteroidonline.com.
$ dig dc-ce20a397.buysteroidonline.com +short
193,111,62,68
Do the "regular Passive DNS dance" from that point...
$ dnsdb query -i 193.111.62.68 -p json | jq -r .rrname |
2nd-level-dom | sort -u
buysteroidonline.com.
flex-lab.de.
planetsteroids.com.
proflexsteroids.com.
server4site.com.
```



# II-d. Working A Kelihos Botnet-Related Spam Example

Leveraging Common Anomalous Text; Expanding And Condensing Hits You Find



#### **Email From Kelihos-Botted Hosts**

- A Farsight staffer like many people -- received unsolicited emails for the pillz host europe-pharm.com on a personal email account
  - The botted hosts sending these spam all appear to have been infected with the spam sending bot known as "Kelihos"
  - Kelihos is the top ranked spambot in the world according to malware experts at McAfee (a unit of Intel).



#### **Unique Creative Text on The E-P Site**

http://www.europe-pharm.com/EN/UK/FAQ#r says:

"In case your order is delayed at customs, *they inform you of that.* They ask the recipient to come and give them *a permission* to open the parcel." [emphasis added]

www.europe-pharm.com is currently at 186.2.163.47

Googling for that odd exact text from the FAQ, we find a number of other sites, including:

https://www.pharmatheke-europe.com/en/faq.html (85.159.236.146)

### **IPs Seen Used By Both Base Domains?**

#### Build a list of IPs used by \*.europe-pharm.com

```
$ dnsdb_query.py -r \*.europe-pharm.com | grep -v
";;" | grep -v "^$" | awk '{print $4}' | grep -v
"[a-zA-Z]" | grep "\." | sort -u > x1.txt
```

#### Build a list of IPs used by \*.pharmatheke-europe.com

```
$ dnsdb_query.py -r \*.pharmatheke-europe.com |
grep -v ";;" | grep -v "^$" | awk '{print $4}' |
grep -v "[a-zA-Z]" | grep "\." | sort -u > x2.txt
```

#### Keep the IPs Common to Both

```
$ comm -1 -2 x1.txt x2.txt > both-x.txt
```

```
$ more both-x.txt
96.45.82.16
96.45.82.201
96.45.83.121
96.45.83.199
```



#### **Base Domains On Each of Those 4 IPs?**

```
$ dnsdb query.py -i 96.45.82.16 | grep -v ";;" | grep -v "^$" |
awk '{print $1}' | 2nd-level-dom | sort -u > y1.txt
$ dnsdb query.py -i 96.45.82.201 | grep -v ";;" | grep -v "^$" |
awk '{print $1}' | 2nd-level-dom | sort -u > y2.txt
$ dnsdb query.py -i 96.45.83.121 | grep -v ";;" | grep -v "^$" |
awk '{print $1}' | 2nd-level-dom | sort -u > y3.txt
$ dnsdb query.py -i 96.45.83.199 | grep -v ";;" | grep -v "^$" |
awk '{print $1}' | 2nd-level-dom | sort -u > y4.txt
$ wc -l y1.txt y2.txt y3.txt y4.txt
     734 y1.txt ← too many!
     663 y2.txt ← too many!
     527 y3.txt ← too many!
     475 y4.txt ← too many!
    2399 total
$ comm -1 -2 y1.txt y2.txt > phase1.txt
$ comm -1 -2 phase1.txt y3.txt > phase2.txt
$ comm -1 -2 phase2.txt y4.txt > phase3.txt
$ wc -1 phase3.txt
    39 ← much better!
$ cat phase3.txt
24livraison-pharmacie.com.
apothekedeutschland.com.
[etc]
```



### The Previous Process, Shown Graphically

pharmatheke-europe.com. www.europe-pharm.com. 109 123 101 244 104.31.80.151 109.123.115.224 104.31.81.151 109.123.120.216 186.2.163.47 109.123.121.248 31.170.178.179 109.123.78.210 31, 170, 179, 179 148.251.91.86 62.212.85.90 217.23.6.184 96.45.82.16 85.159.236.146 96.45.82.201 96.45.82.16 96.45.83.121 96 45 82 201 96.45.83.199 96.45.83.121 96 45 83 199

24livraison-pharmacie.com. apothekedeutschland.com. buonafarmacia.com. casa-farmacia.com. comprar-viagra-sin-receta.net. dietplans10.com. erectionsexuelle com. europe-pharm.com. farmacia-barata-online.com. farmacia-internacio.com. grand-pharmacie.com. quteapotheke.com. guteapotheke.net. lekarna-cz.com. levitraexpress.com. lida-es.com. meddirekt24.com. meddirekt24.org. medicamentfrance.net. mueller-apotheek.com.

pharmacie-en-ligne-france.com. pharmacieviagra.net. pharmatheke-europe.com. potenzmittelonline24.com. potenzpillenversand.com. potenztheke com. premium-pharm.com. priznakynemoci.com. ratiopharma.net. samplecialis com. slim-med24.com. super-potenzmittel.com. viagra-kaufen-apotheke.com. viagra-pillen-kaufen.com. viagra-pour-femme-fr.com. viagra-prodej-bez-receptu.com. viagrakupno.com. vkaufen.com. wo-cialis-bestellen.com.



# III. Going "Multivariate"

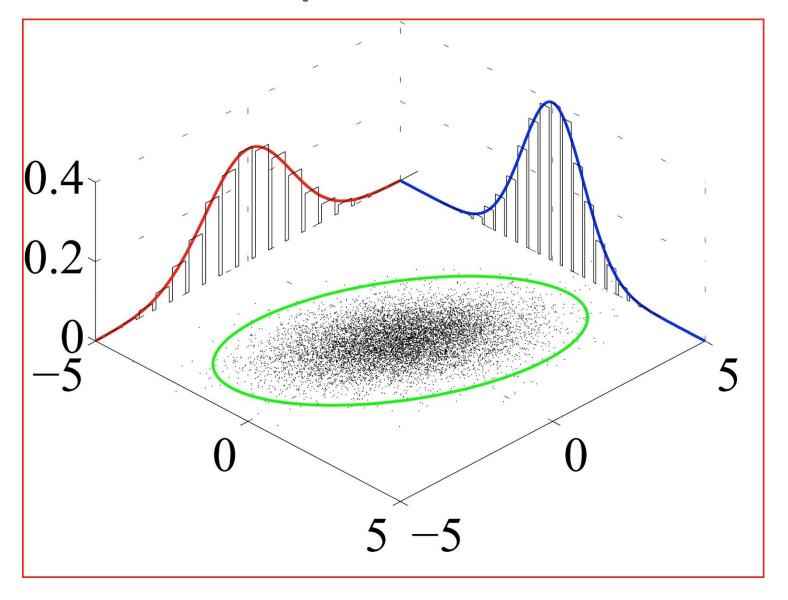


## Points In An n-Dimensional Space

- In a multivariate approach we look at more than one measurement at the same time
- This allows "interactions" to be accounted for:
  - x by itself? okay
  - y by itself? okay
  - x and y combined together? Kablooey! (online equivalent of tranquilizers taken with cocktails)
- NOT combining multiple attributes into a single score, compared against a threshold (SpamAssassin style)
- NOT just successive application of independent univariate filters, either



# A Simple Two-D Normal Distribution





#### The Data We've Got

- Currently passive DNS captures data about three main types of DNS-related entities:
  - Names
  - IPs
  - Name servers
- None of that is beautiful continuous data.
- If you attempt to visualize it, it will NOT look like the pretty graph on the preceding page.



#### **Review: Level of Measurement**

- We prefer discrete measures (integer counts) to just binary attributes (present/absent)
- We prefer continuous measures (real decimal numbers) to just discrete measures (integer counts)
- Visualized, non-continuous measures turn into "boxy" spaces where "clumps" of observations land in one cell or another of a table rather than forming a smooth ndimensional density distribution. You can try skyscraper charts, but they're ugly.



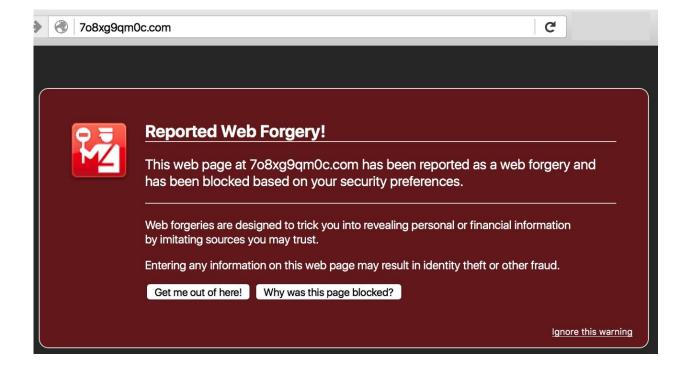
## **Caching Impacts**

- Passive DNS data may also include timing data ("domain age"), implicit/approximate volume data ("domain usage"), and of course we also get to see the values of the domain names themselves
- **Unfortuately, DNS caching** partially interferes with potential timing/volumetric analyses (with the longer the TTL, the worse the obfuscation becomes).
- Sometimes, differences are so unmistakable that even caching and TTLs effects don't matter.



# FARSIGHT Volume Data (Notwithstanding Caching)

- Easy: which domain is less well established / less trustworthy?
  - \$ dnsdb query.py -r www.google.com/a | grep count | awk '{print \$3}' | sed 's/,//g' | paste -sd+ - | bc 1795747251 ← observations we've seen...
  - \$ dnsdb query.py -r 708xg9qm0c.com/a | grep count | awk '{print \$3}' | sed 's/,//g' | paste -sd+ - | bc 1109 ← observations we've seen
- Confirmed →

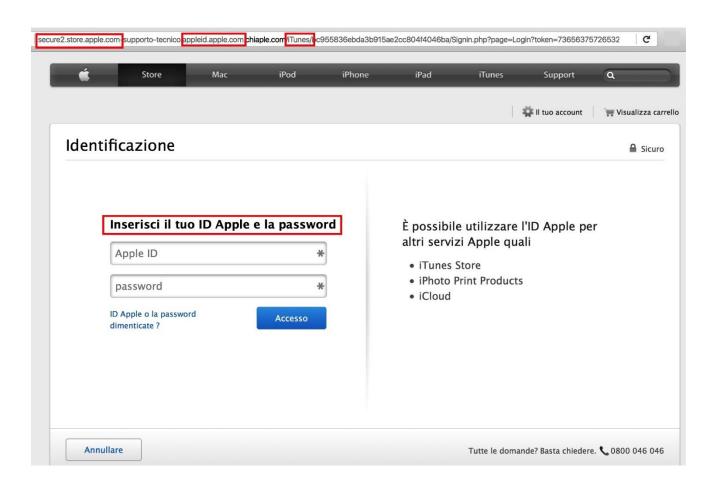




### **Some Hostnames Speak For Themselves**

Anyone ready to buy a new Mac? (Don't get <u>phished!</u>)

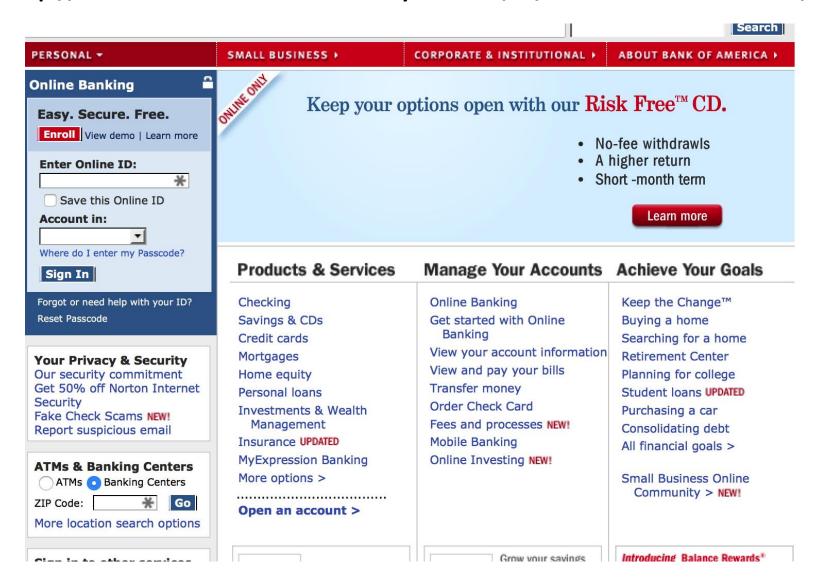
hxxp://secure2.store.apple.com-supporto-tecnico.appleid.apple.com.chiaple.com/





# A Is For Apple, B Is For BOA, ....

hxxp://bankofamerica.com.bosnaknakliyat.com.tr/us/www.bankofamerica.com/



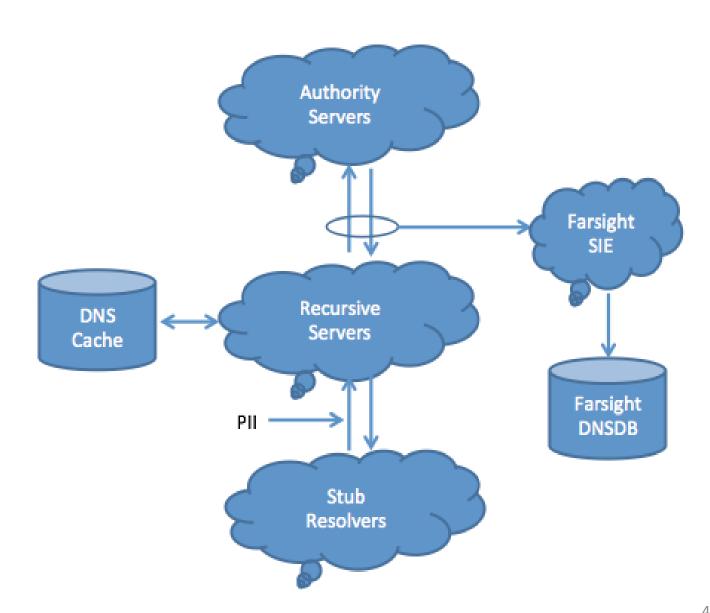


## **Exercising Self-Restraint**

- Many other measurable passive DNS characteristics are intentionally NOT collected
- This means:
  - No ultimate end-user query source IP
  - No "query stream of successive queries" associated with just a specific unique user
  - No sensor identity/location data
  - Etc.



# **Collecting Above The Recursive**





# **Augmenting Classic pDNS**

- Combine Passive DNS data with other non-DNS data to "go multivariate."
  - Non-DNS data could be pre-existing data such as domain Whois or IP whois data.
- Collect new data to augment passive DNS dataset (where active scanning is allowed by law and by your network terms of service).
  - For example, fingerprint/scan hosts with NMAP or a similar scanning tool to see what pattern of ports (if any) are open on a range of IPs.



# Loose or "Fuzzy" Matches

- Not a new concept; a few common examples you may be familiar with
  - agrep (various algorithms used, see http://www.tgries.de/agrep/#ALGORITHMS)
  - "similar sounding string" matching (Soundex, Metaphone [2,3], etc.)
- Should "loose matches" conceptually also include "a network neighborhood around" a starting IP (e.g., "in the same netblock," "routed by the same ASN," "within a span of N intervening IP addresses," etc.)?



#### **Conclusion**

- Passive DNS is a highly effective tool to enrich threat intelligence and advance digital investigations
- As Passive DNS grows in use and popularity, users are finding new ways to use the data as well as finding possible roadblocks
- Bad Guys may obfuscate their digital trail to make connecting the dots of the investigation more difficult
- There are a number of work-around techniques you can use around obfuscation including augmenting Passive DNS with new or existing data such as domain Whois or IP Whois.
- Understanding the many uses of Passive DNS can help your organization stay ahead of the threat