Multivariate Solutions to Emerging Passive DNS Challenges

Dr. Paul Vixie, CEO and Dr. Joe St Sauver, Scientist
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
[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 3rd party domains

- Many related domains use the same set of dedicated name servers, with no innocent 3rd 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 Doesn't 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 comingled inextricably with innocent 3rd party resources.

• Bad Guy “Hit and run" scenarios
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.
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.
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 terrific target by law enforcement
"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 3rd 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
"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
'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?"

![Website Screenshot](image)

**About Us**

We are a business-to-business (B2B) and business-to-customer (B2C) online factory, which is a comprehensive trade solution provider and offers one-stop trade services to international buyers and wholesalers.

Customer's satisfaction is our top concern and we will do our best to exceed your expectations by constantly introducing exceptional quality goods, prices and world-class customer service.

We assure you of reliable quotations, prompt deliveries and stable supplies. All our goods are carefully packaged and delivered worldwide within 3-8 days.

Shopping online, you need not to pay any shipping or tax, therefore, we would give you any invoice. Our factory also has one of the most active and enthusiastic communities on the net.

Shoppers are welcome to communicate with us and tell us their needs. The information of our web can be shared on blogs or emailed to friends. Our enthusiastic groups provide ideas on various shopping and offer shoppers plenty of choice to choose.

Thank you for visiting our website and we hope you enjoy your shopping experience with us.

---

©2009-2016 official49ersjerseys.com. All Rights Reserved.

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Create Date</th>
<th>Registrar</th>
<th>Registrant Name</th>
<th>Registrant Organization</th>
<th>Registrant Street</th>
<th>Registrant City</th>
<th>Registrant State/Province</th>
<th>Registrant Postal Code</th>
<th>Registrant Country</th>
<th>Registrant Phone</th>
<th>Registrant Fax</th>
<th>Registrant Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>official49ersjerseys.com</td>
<td>2015-09-03 14:24:36</td>
<td>SHANGHAI MEICHENG TECHNOLOGY INFORMATION DEVELOPMENT CO., LTD.</td>
<td>shao nian</td>
<td>shao nian</td>
<td>Shang Hai Shi Qu</td>
<td>shanghaishi</td>
<td>shanghai</td>
<td>123123</td>
<td>CN</td>
<td>+86.0211231231</td>
<td>+86.0211231231</td>
<td><a href="mailto:cj2015tit@126.com">cj2015tit@126.com</a></td>
</tr>
<tr>
<td>nflshop.com</td>
<td>2015-07-14T04:00:24-0700</td>
<td>MarkMonitor, Inc.</td>
<td>NFL Enterprises LLC</td>
<td>NFL Enterprises LLC</td>
<td>345 Park Ave.,</td>
<td>new york</td>
<td>ny</td>
<td>10017</td>
<td>US</td>
<td>+1.2124502000</td>
<td>+86.0211231231</td>
<td><a href="mailto:dns_admin@nfl.com">dns_admin@nfl.com</a></td>
</tr>
</tbody>
</table>

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
104.27.143.198  ➔ Hidden behind Cloudflare
104.27.142.198  ➔ Hidden behind Cloudflare

$ dig official49ersjerseys.com mx +short
0  dc-96d9f219.official49ersjerseys.com.

$ dig dc-96d9f219.official49ersjerseys.com +short
107.155.198.200  ➔ NOT hidden behind Cloudflare (Sentris)

Do the "regular Passive DNS dance" from that point...

$ dnsdb_query -i 107.155.198.200 -p json | jq -r .rrname | sort -u
cheapcustomjerseysonline.com.
dallascowboymall.com.
dallascowboysmall.com.
[etc]

dnsdb_query (c lang)? see https://github.com/dnsdb/dnsdb_c

Get jq from https://stedolan.github.io/jq/
#!/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

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Create Date</th>
<th>Registrar</th>
</tr>
</thead>
<tbody>
<tr>
<td>cowboysonlinestore.com</td>
<td>2015-09-16</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialbroncosstore.com</td>
<td>2015-11-06</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialredskinsjersey.com</td>
<td>2015-09-04</td>
<td>cndns.com</td>
</tr>
<tr>
<td>falconsonlineshop.com</td>
<td>2015-09-16</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialeaglesshop.com</td>
<td>2015-11-09</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialpackersstore.com</td>
<td>2015-11-05</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialsaintsstore.com</td>
<td>2015-08-14</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialsaintsstore.com</td>
<td>2015-08-14</td>
<td>cndns.com</td>
</tr>
<tr>
<td>officialeaglesstore.com</td>
<td>2015-11-09</td>
<td>cndns.com</td>
</tr>
</tbody>
</table>
II-c. Overcoming Obfuscation:

Scheduled Controlled Substances

Illustrates Use of TXT Record Info To Overcome Reverse Proxy Usage
Anabolic Steroids Are Schedule III

<table>
<thead>
<tr>
<th>SUBSTANCE</th>
<th>DEA NUMBER</th>
<th>CSA SCH</th>
<th>NARC</th>
<th>OTHER NAMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-methylfentanyl</td>
<td>9814</td>
<td>I</td>
<td>Y</td>
<td>China White, fentanyl</td>
</tr>
<tr>
<td>Alpha-methylthiofentanyl</td>
<td>9832</td>
<td>I</td>
<td>Y</td>
<td>China White, fentanyl</td>
</tr>
<tr>
<td>Alpha-methyltryptamine</td>
<td>7432</td>
<td>I</td>
<td>N</td>
<td>AMT (Positional Isomer: N-Methyltryptamine)</td>
</tr>
<tr>
<td>Alphaprodine</td>
<td>9010</td>
<td>II</td>
<td>Y</td>
<td>Nisentil</td>
</tr>
<tr>
<td>alpha-pyrrolidinobutiophenone (α-PBP)</td>
<td>7546</td>
<td>I</td>
<td>N</td>
<td>1-phenyl-2-(pyrrolidin-1-yl)butan-1-one</td>
</tr>
<tr>
<td>alpha-pyrrolidinopentiophenone (α-PVP)</td>
<td>7545</td>
<td>I</td>
<td>N</td>
<td>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)</td>
</tr>
<tr>
<td>Alprazolam</td>
<td>2882</td>
<td>IV</td>
<td>N</td>
<td>Xanax</td>
</tr>
<tr>
<td>AM-2201 (1-(5-Fluoropentyl)-3-(1-naphthoyl) indole)</td>
<td>7201</td>
<td>I</td>
<td>N</td>
<td>AM-2201</td>
</tr>
<tr>
<td>AM-694 (1-(5-Fluoropentyl)-3-(2-iodobenzoyl) indole)</td>
<td>7694</td>
<td>I</td>
<td>N</td>
<td>AM-694</td>
</tr>
<tr>
<td>Aminorex</td>
<td>1585</td>
<td>I</td>
<td>N</td>
<td>has been sold as methamphetamine</td>
</tr>
<tr>
<td>Amobarbital</td>
<td>2125</td>
<td>II</td>
<td>N</td>
<td>Amytal, Tuinal</td>
</tr>
<tr>
<td>Amobarbital &amp; noncontrolled active ingred.</td>
<td>2126</td>
<td>III</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Amobarbital suppository dosage form</td>
<td>2126</td>
<td>III</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Amphetamine</td>
<td>1100</td>
<td>II</td>
<td>N</td>
<td>Dexedrine, Adderal, Obetrol</td>
</tr>
<tr>
<td>Anabolic steroids</td>
<td>4000</td>
<td>III</td>
<td>N</td>
<td>&quot;Body Building&quot; drugs</td>
</tr>
</tbody>
</table>

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\(^{st}\) Offense: $500,000/$2.5 million Up To 15 years

2\(^{nd}\) 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).
"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]
**IPs Seen Used By Both Base Domains?**

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

```bash
$ 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**

```bash
$ 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**

```bash
$ 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
$ 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 -l phase3.txt
  39 ← much better!
$ cat phase3.txt
24livraison-pharmacie.com.
apothekedeutschland.com.
[etc]
The Previous Process, Shown Graphically

104.31.80.151
104.31.81.151
186.2.163.47
31.170.178.179
31.170.179.179
62.212.85.90
96.45.82.16
96.45.82.201
96.45.83.121
96.45.83.199

pharmatheke-europe.com.
109.123.101.244
109.123.115.224
109.123.120.216
109.123.121.248
109.123.78.210
148.251.91.86
217.23.6.184
85.159.236.146
96.45.82.16
96.45.82.201
96.45.83.121
96.45.83.199

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

pharmacieviegra.net.
pharmatheke-europe.com.
potenzpille-versand.com.
potenztheke.com.
premium-pharm.com.
priznakynamoci.com.
ratiopharma.net.
samplecialis.com.
super-potenzmittel.com.
viagrakupno.com.
vakupno.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? Kablooe! (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

https://commons.wikimedia.org/wiki/File:Multivariate_normal_sample.svg
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.
Statistical options for nominal data are limited: you can do cross tabs, but (a) that's not very statistically "sexy," and (b) interpretation becomes hard as the table size increases.
Other Data? Volume (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 7o8xg9qm0c.com/a | grep count | awk '{print $3}' | sed 's/,//g' | paste -sd+ - | bc
    1109 ← observations we've seen

• Confirmed→

Reported Web Forgery!

This web page at 7o8xg9qm0c.com has been reported as a web forgery and has been blocked based on your security preferences.

Web forgeries are designed to trick you into revealing personal or financial information by imitating sources you may trust.

Entering any information on this web page may result in identity theft or other fraud.

Get me out of here!   Why was this page blocked?

Ignore this warning
Some Hostnames Speak For Themselves

• Anyone ready to buy a new Mac? (Don't get *phished*)

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

Diagram showing a network of DNS servers and databases with interactions labeled as PII.
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.
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.
Thank you!

Dr. Paul Vixie
vixie@fsi.io