Passive DNS And The Halting Problem

Joe St Sauver, Ph.D. (stsauver@fsi.io)
Distributed System Scientist
Farsight Security, Inc.

B|Sides Vancouver, BC, March 2015

https://www.stsauver.com/joe/halting-problem/

Disclaimer: Opinions expressed are solely those of the author.

I. Introduction

Our Format Today

- I'm glad we're in Room II, because it means that we can have a more interactive format than would be possible in the larger room here at the Imperial.

- I've put together some material I'm planning to go over, but I encourage you to ask questions as we go along, should questions arise – I've explicitly left some time to do so.

- If you don't like asking questions in front of the whole group, I'm also happy to chat after the session, or to visit by email

- These slides will be available on the B|Sides Vancouver web site, and from my web site, https://www.stsauver.com/joe/

My Odd Slide Style

- Speaking of slides... Let me get it "out of the way" up front: I produce **detailed slides.** This style drives some people **crazy,** so I like to explain why I do things this way.
- I've tried the more-typical 3-4 bullets/slide (with ~15 slides for an hour long talk), but I find myself getting **sidetracked, rambling/ running over,** or I end up **missing/skipping stuff.**
- I also deal with **complex issues**, and I HATE to be misquoted.
- My slide style prevents a lot of those problems, and means that **you don't need to try to take notes.**
- That said, I'm **not going to read my slides word-for-word for you.** You don't need to try to do so, either, although they are a sort of "closed captioning" if you're deaf or hard-of-hearing.
- I also write detailed slides to help people looking at them after the fact, and for indexing by web search engines.

My Background

- My Ph.D. is in Production and Operations Management
- Spent ~28 yrs at the Univ of Oregon Computing Center in Eugene
  - Worked under contract for Internet2 (the U.S. counterpart to Canarie.CA) as their Nationwide Security Programs Manager under contract through UO, including running the highly successful InCommon SSL/TLS Certificate Program and the InCommon Multifactor Program
- I left UO in 4Q14 to go to work for Internet Hall of Fame award recipient Paul Vixie and his new company, Farsight Security, Inc.
- I remain active in a variety of industry-related roles including:
  - Senior Technical Advisor, M3AAWG (I also lead the Identity Management SIG, and am particularly active in the anti-Pervasive Monitoring SIG)
  - I remain on the Technical Advisory Group of the REN-ISAC
  - I'm a Strategic Advisor, Online Trust Alliance (OTA)
  - Board Member, Coalition Against Unsolicited Commercial Email (CAUCE)
  - Community Representative, Broadband Internet Technical Advisory Group

II. Our Discussion Today

Spam and Canada's Anti-Spam Legislation (CASL)

- Let me take a moment to congratulate all Canadians on passing an excellent anti-spam law. It was a happy day when CASL came into effect on January 15th! If you've not had a chance to review it yet, you should! Look at http://www.fightspam.gc.ca/
- I'm particularly impressed to see the CRTC off to a fast start with a $1.1 million dollar penalty imposed just this March 5th. Nice work CRTC peopl!
- I'm hoping that all Canadians enjoy improved protection from unwanted email now, and well into the future.
- Seeing the terrific strides that Canada has taken also made me want to talk about using passive DNS to research cyber abuse "at scale," particularly when it comes to unsolicited email.
- I also wanted to talk about how to figure out when it's time to stop.

The Halting Problem

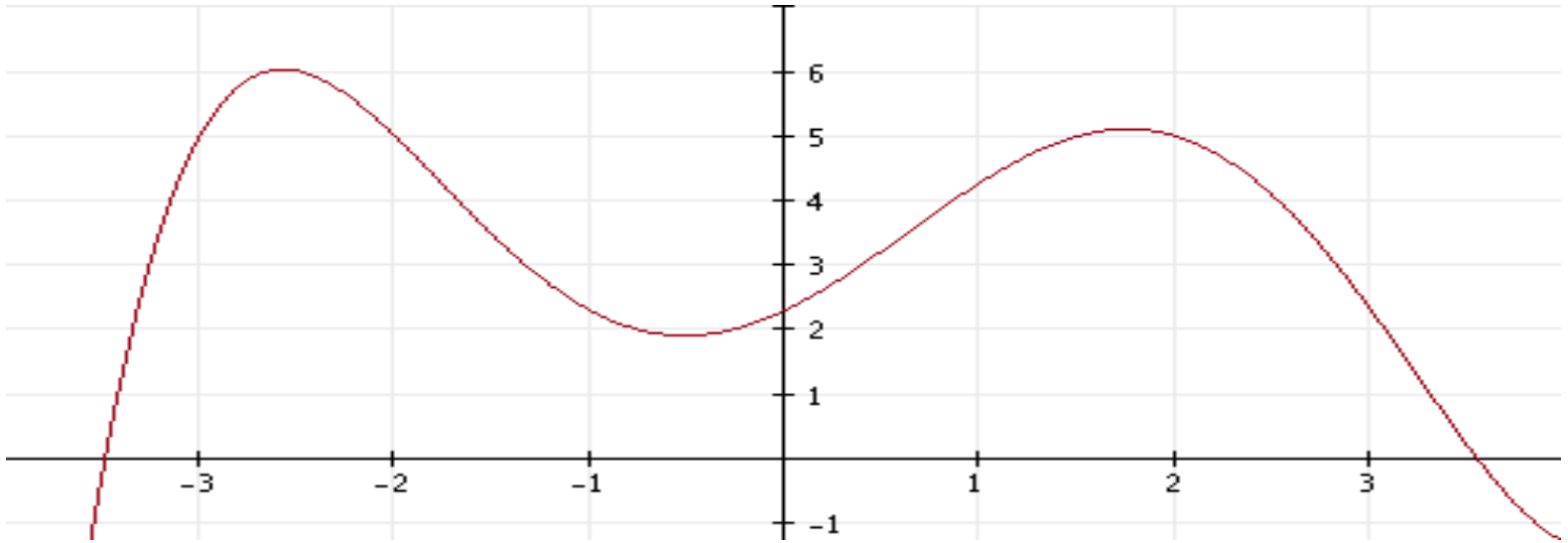
- You might not expect it to be hard to figure out when to quit: common sense says do the job, and when you're done, that's it.

- Sometimes that's exactly right. There are some problems that have a finite series of steps, and once you've completed those steps, there's an answer, and you're done. ("Count the number of candies in this bag.")

- Other problems are special because you may never know if you're "done," or if you've found the "best" answer, or if you've found the "complete" answer.

- Consider the graph on the next slide.

Numerically find max f(x) for x in (-∞, +∞)



- If we (naively) start at x=0, and then use the simple (bad) heuristic "go either -Δ left or +Δ right, as may most increase f(x), until f(x) begins to decrease," we'll eventually end up at the top of the right hand hill at x=~1.75, y=~5.1. That's a local maxima.
- But what about x=~-2.5, y=~6, on the left? It's obviously higher.
- And what about values you don't see, "off screen?" How do we know that there isn't a better solution out there "somewhere?"

You Could Try Random Starting Points, Different Δ's

- We can try to avoid hitting a local maxima by picking random starting points (rather than just starting at x=0), or by picking different Δ's, but we're still just numerically "rolling the dice" when it comes to finding a global (rather than local) maxima.
- You might now say, "Joe! How does this obscure mathematical thing apply to cyber security?"
- The answer is that when you're chasing clues in network forensics, you face a similar problem: how do you know you've found "everything" that's relevant? How do you know you're "DONE?"

- **ACTIONS TARGETING VARIOUS UNWANTED ONLINE BEHAVIORS ARE OFTEN NOT DONE AT SCALE, AND IN FACT ARE ALL-TOO-OFTEN GROSSLY INCOMPLETE.**

If Only Cyber Criminals Would Politely and Cooperatively Create Simple Infrastructures!

- Imagine a bad guy/bad gal using:
  -- Just **one domain name** (with accurate and un-privacy/proxy protected domain whois information)
  -- No IP agility, just **one never-changing static IP address**
  -- Name servers in the same domain and on the same IP
  -- Static, publicly accessible evil content, conveniently indexed by major search engines

- Straightfoward to investigate, take down, or prosecute.

- But of course, that's NOT what cyber criminals do. They want to avoid detection and takedowns/prosecutions, so they like to employ more complicated architectures, instead.

"Utopia" for the Bad Folks

- **A portfolio of many seemingly normal/routine domain names to use at the same time**, or
- **Constantly changing domain names** (in some cases used for just minutes before being dumped and replaced).
- **One IP address now, different IP addresses later**, or
- **Many IP's in use at the same time** (sometimes even fast fluxing).
- **Name servers that may also be fast flux,** or
- **Shared name servers** that are shared across tens of thousands of other domains, a few bad, but many good)
- **Domain whois data that is privacy protected**, and **IP whois data that doesn't show any reassignments** (no SWIPs/rwhois)
- **Content display that varies** according to visiting IP, the referrer string, the number of previous visits, etc.

What About the Good Folk? Routine "Stretch" Goals

- **Find "ALL" closely-related evil IPs**
  - This helps identify compromised machines that may need remediation,
  - It avoids leaving the miscreant with a base from which to recover
  - You maximize chance of successfully chasing financial payment details and other business records for things like command and control servers
  - You may even discover additional unknown criminal lines of business

- **Find "ALL" closely-related evil domains**
  - Cyber criminals are using more and more domain names, so no choice
  - More domains seized = more "news worthy" action, and more incentive to spend limited cycles on *this* case, not other alternative ones
  - If you're an LEO, you wouldn't want to end up with an incomplete takedown/seizure (you know, potentially spawning online remarks such as "Hah hah hah, they seized a dozen of my domains, but they missed two hundred others, so I didn't even really notice it")

Domain Takedowns/Seizures

- It's not my plan to get into a discussion or whether domain takedowns/seizures are a "good" or "bad" tool for use by civil and/or criminal authorities. I've seen them used very carefully and professionally and effectively in ways that minimize or eliminate collateral damage, and I've seen them used in ways that seem to an outsider both ineffective and reckless.

- Today, however, I'm simply going to note that most takedowns are ***not* exhaustive**, and many **cannot *be* exhaustive** (a miscreant can always register another domain)

- Consider the March 2012 seizure of bodog.com, still returning a "seized domain" parking page today. However, bodog's brand remains up and operational in other TLDs, right?

Another Take Down Focus Area: Counterfeit Goods

- An easy-to-see example of the *growth* in the volume of domains seized/year can be seen in actions related to counterfeit goods:
  - 82 domains seized Nov 2010 http://www.ice.gov/doclib/news/releases/2010/domain_names.pdf
  - 132 domains seized Nov 2011 http://www.ice.gov/news/releases/ice-european-law-enforcement-agencies-and-europol-seize-132-domain-names-selling
  - 307 domains seized Feb 2012 http://www.wired.com/2012/02/sports-domains-seized/
  - 690 domains seized Dec 2013 https://www.europol.europa.eu/content/690-internet-domain-names-seized-because-fraudulent-practices
  - Nearly 5,000 domains allegedly seized under seal in October 2014 http://www.internetcommerce.org/undue-process/
- But even with that rough doubling year over year, the authorities really still aren't executing domain seizures **"at industrial scale"**

Example Of What I Mean By "*At Scale:*" Conficker

- If you really want to see what I think of when I think of DNS takedown operations "at scale," consider something more like the 2nd major version of Conficker. As Farsight CEO Paul Vixie noted in his US Senate Judiciary Committee testimony from last summer:

  Conficker's second major version generated 50,000 (fifty thousand) domain names per day that had to be laboriously blocked or registered in order to keep the control of this botnet out of the hands of its criminal authors. Complicating the situation, these 50,000 domain names were split up across 110 different "country code" top-level domains that are each the property of a sovereign nation. The registries for these domains are a mix of private and public institutions, some with national government oversight and many without. Almost all of the 110 registries agreed to cooperate [...]

  http://www.judiciary.senate.gov/imo/media/doc/07-15-14VixieTestimony.pdf

It's Rare to See Examples of Action "At Scale"

- Some miscreants have evolved a survival strategy that emphasizes staying **"too small to bother taking down"**

- Other miscreants have learned that they can ALSO avoid takedowns simply by **out scaling investigators**
  - They can create a new domain for a few bucks in mere seconds
  - The authorities have to devote days or hours (and substantial expense) to trying to knock those domains down

  This was true for SOPA/PIPA (see https://www.stsauver.com/joe/managers-amendment/sopa-amended-version.pdf , when prompted "final" without the quotes), and it's true for most other take-down-oriented approaches today, except for voluntary use of blocklists (but Uncle Sam and Johnny Canuck don't publish blocklists!)

Only a Few Cyber Crime Areas Are Universally Hated

- Large-scale malware (like Conficker) is one such area. It's hated because it is so disruptive, and because large scale DDoS attacks rely on botted hosts. Providers get that, and that's why they voluntarily collaborate to deal with this threat.
- Child abuse materials are another obvious example of online conduct that everyone agrees is totally unacceptable. Law enforcement officers are willing to work long hours, and across borders, to stamp it out.
- Phishing is a third widely-regard-as-unacceptable online activity.
- But email spam? Web spam? Sadly, all too many are still unwilling to make the effort to mitigate it. It has been more or less a "safe crime" for many miscreants, at least pre-CASL.
- I'm hoping to see Canada lead the charge against unsolicited email, ***at scale***

Enabling Action Against Unsolicited Email At Scale

- We need to help investigators understand how to use tools that WILL work at scale.

- More investigators need to learn how to start with one "thread" or "clue," using those starting points to uncover connections that can then lead to other threads that can also be successfully chased down

- The best tool I can think of for doing this is **passive DNS.**

III. DNS vs. Passive DNS

A Few Limitations of Regular DNS

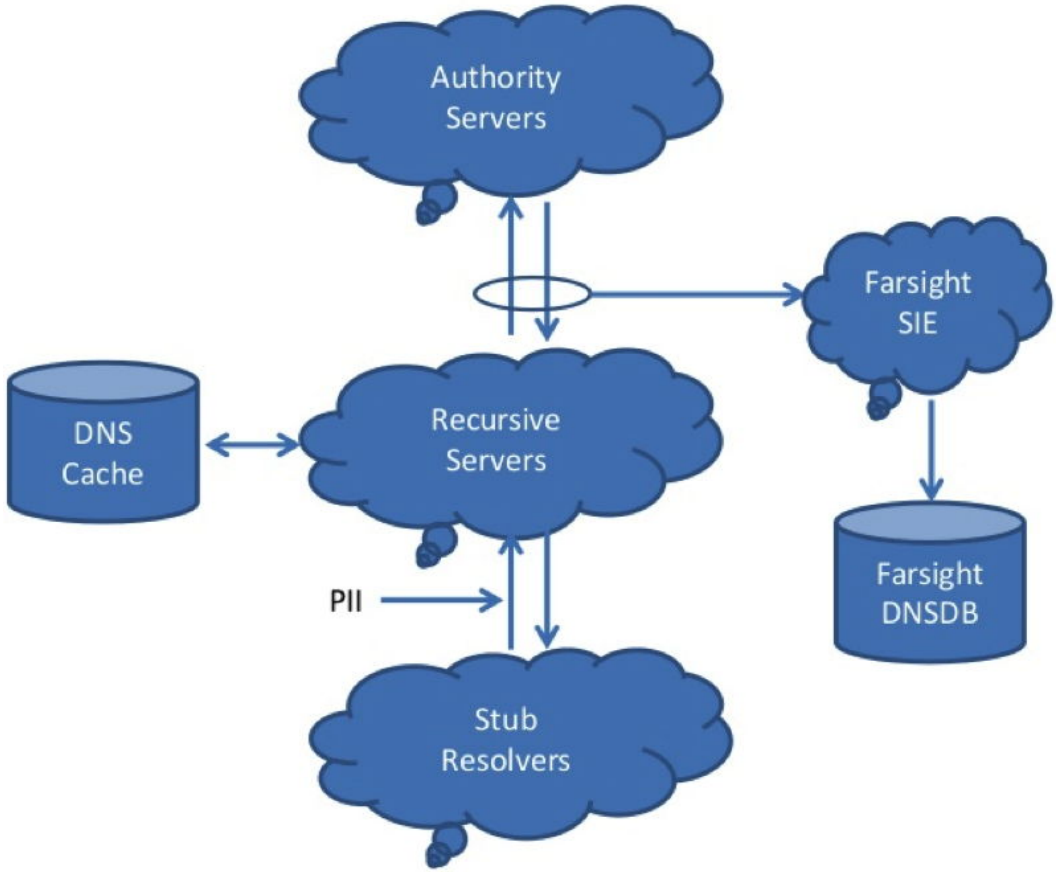
- Normal DNS is all about ***what's defined NOW***, not what was true last week or last month when an incident may have happened
  - Caching can complicate interpretation
  - No way to "time shift" just to a period of interest
  - No historical record of what's happened earlier
- Regular DNS assumes authoritative DNS has been ***properly configured*** to return consistent (and non-deceptive) answers
  - Content distribution networks may tailor results by query location
  - To be able to get in-addr queries, PTR records must have been created
  - Need to hope that name servers haven't been directed to selectively respond with bad data (e.g., via use of RPZ) for some query sources
- Regular DNS works best for ***simple architectures***.
  - Even basic igh density virtual hosting environments can be tricky: with perhaps five hundred domains hosted on one IP – which (if any) of those 500 domains should be returned as a PTR?

Some Infeasible Regular DNS Queries

- Show me *all* FQDNs under example.com (e.g., *.example.com)
- List all the domains that use the name server ns1.example.com
- Given the IP used by one FQDN, what other domain names are on that same IP address?
- Given a domain name, does it resolve to multiple (different) IPs? Has it moved around over time? If so, to what IP addresses?
- Show me all the SMTP TLSA domains (e.g., all domains matching _25._tcp.*)
- Which new gTLDs appear to be most popular?
- What domains have been used in the last five minutes?
- These queries are examples of the sort of things that "regular DNS" was just NOT designed to address...

Passive DNS Collection

- Passive DNS data is collected by passively monitoring DNS cache miss traffic above recursive resolvers, and then saving and indexing that data in a specialized database
- This allows a wide range of otherwise-impossible queries to be easily done



- And best of all, your investigative target will not know that you're "on the hunt" – they won't be able to see your queries

An Aside About Passive DNS and Privacy

- I care a great deal about privacy, and I hope you all do, too.
- At the same time, I want to ensure that cyber criminals aren't allowed to operate with impunity online. Law breakers must be held accountable for their crimes, online or in physical space.
- Passive DNS, collected properly, comes from above large shared recursive resolvers. Queries appear to originate from the recursive resolver, not any individual user. As a result, no personally identifiable information is collected or stored.
- Because of this architecture, passive DNS does NOT raise the sort of pervasive monitoring concerns that are widely associated with things like bulk metadata collection and traditional traffic analytic methods, as discussed in "The Enduring Challenges of Traffic Analysis," https://www.stsauver.com/joe/flocon-2015/

Various Organizations Offer Passive DNS

- DNSDB from Farsight (https://www.farsightsecurity.com/)
  - FSI was spun off from ISC; DNSDB is a commercial product offered by FSI.
  - Individual law enforcement officers, academic researchers, and non-profit orgs can still request full or partially underwritten access to DNSDB.
  - Run a sensor and contribute data? That will also be financially recognized.
- Other passive DNS implementations:
  - Florian Weimer's BFK, http://www.bfk.de/bfk_dnslogger.html
  - CERT.at/Aconet Passive DNS (inquire: kaplan@cert.at or lendl@cert.at)
  - CIRCL Passive DNS, http://www.circl.lu/services/passive-dns/
  - http://passivedns.mnemonic.no/search/
  - https://www.opendns.com/enterprise-security/resources/data-sheets/investigate/
    https://www.cs.auckland.ac.nz/research/groups/sde/dhdb-index.php
  - VirusTotal, https://www.virustotal.com/#search
  - 360.cn Passive DNS, https://www.passivedns.cn/help/

Access to Passive DNS Is Typically Vetted

- Passive DNS providers are typically quite careful when it comes to permitting access to passive DNS resources – no one wants passive DNS to be used for malicious purposes such as reconnoitering a site targeted for a malicious attack.

- Don't be surprised or offended if people want to know who you are, how you plan to use passive DNS, and who can vouch for your *bona fides*.

- It's appropriate for powerful tools to be carefully controlled.

IV. Using Passive DNS

Passive DNS Is All About Making Connections

- There are hugely useful relationships – links – that are defined in the domain name system.

- Passive DNS is all about collecting DNS data, saving it in a highly optimized data, and then efficiently searching that data for those otherwise often overlooked relationships.

Passive DNS Is Not About Goodness or Badness

- There's a temptation to assume that all cyber security data must be about "goodness" or "badness" these days.

- That's not the case for passive DNS. Passive DNS is about objective connections between domains and IPs, and name servers and domains. (It's like GeoIP data in its objectivity)

- Passive DNS is NOT about goodness or badness, just connections.

- **You**, however, may choose to draw inferences about things that are in the same neighborhood as known badness, particularly if the things you find look very similar to in-hand examples of known badness.

BASIC PASSIVE DNS STRATEGY

- **If you've got IP addresses**, look for domain names associated with those IPs (if you're using Farsight's dnsdb_query command line client, this is a dash i query)

- **If you've got domain names**, look for IP addresses associated with those domains names (this is a dash r query)

- **If you've got name server FQDNs**, look for domain names known to use those name servers (this is a dash n query)

- See https://www.farsightsecurity.com/Blog/20150311-stsauver-rrset-rdata/

- **Note:** Key point -- do NOT run just one query and then STOP!

Some Passive DNS Processing Techniques

- Chain/Pivot: feed output from one step back into the next step
- Condense: reduce wildcarded fully qualified domain names to 2nd level domains (save any name server-related records first!)
- Decouple: take *just* domain names OR *just* IPs from results
- Deduplicate: Use "sort | uniq" commands or the equivalent
- Prioritize: sort in descending order by observed record count
- Enhance: seek additional data (e.g., from Spamhaus, whois, etc.)

Some Fog Lines on the Highway

- Having started from an initial handful of clues, you want to find more related domains: amplify and enhance the data you've already got.

- Avoid miscreant attempts to "blow you out of the water" by exploiting wildcarding, fast flux hosting, etc.

- Don't contaminate a list of bad domains with good ones; minimizing false positives is key. Don't just take everything you find at face value! Be skeptical! Be CAREFUL!

- Execute crisply. The bad guys are in and out quickly. They're not going to wait around for you to dilly-dally.

V. An Unsolicited Email Example

Unsolicited Email

- Unsolicited email can provide us with a nice example to illustrate our point/help demonstrate passive DNS tools.

- I've also got to confess that I have a certain fascination with unsolicited email that stems in part from my work with M3AAWG, CAUCE, and other groups.

- I also like using unsolicited email examples (instead of malware examples) because there's a lower chance of infection if you visit sites I mention (however, I strongly urge you to NOT visit any domain mentioned here, unless you do so in a sandboxed virtual machine – you just can't predict what may happen)

Sample Message Bodies For Today's Discussion

- Consider the following six "message bodies" (none very long), as seen in real email messages:

- http://meetinghouse.hwqdjzbw.eu/ [Canadian Health&Care Mall]
- http://totalitarianism.mymedicinalstore.in/ [ditto]
  The first-rate way to please your girlfriend
- Do you wish to gratify your loved one every night?
  http://behalf.besthealingelement.ru/ [ditto]
- http://telethon.purepharmacygrouponline.ru/ [ditto]
  Very good method to develop your intimate life
- http://fascist.homedrugquality.com/ [ditto]
  Stimulate better enlargement
- http://glowing.onlinecuringbargain.in/

Naive Observations/Questions

- That sure seems like a "lot" of different domain names, all pointing at the same underlying entity, eh?
- Wonder how many other domains are related to these guys? Can we identify ALL of those domains? For example, do they all live on the same IP address?
- And how do we know we've found "all" of them, and can stop? How do we know that there isn't some additional pod of domains out there?

- Let's begin by resolving the domains we know with dig. Maybe they'll "clump" or "cluster" around just a couple of IP addresses?

Resolving Those Domains with dig

- meetinghouse.hwqdjzbw.eu ➔ 95.31.192.232
- totalitarianism.mymedicinalstore.in ➔ 95.31.192.232
- behalf.besthealingelement.ru ➔ 95.31.192.232
- telethon.purepharmacygrouponline.ru ➔ 95.31.192.232
- fascist.homedrugquality.com ➔ 95.31.192.232
- glowing.onlinecuringbargain.in ➔ 95.31.192.232

A little later…

- meetinghouse.hwqdjzbw.eu. ➔ 213.169.149.3
- totalitarianism.mymedicinalstore.in. ➔ 213.169.149.3
- behalf.besthealingelement.ru. ➔ 213.169.149.3
- telethon.purepharmacygrouponline.ru. ➔ 213.169.149.3
- [etc]

Clustering DID Happened...

- **Simply by resolving the domains, we were able to condense our 6 original domain names to 2 IP addresses**

- It seems reasonable to take those two IPs as indicators potentially worthy of further analysis

What Can Whois Tell Us About Those Two IPs?

- **95.31.192.232:**
  inetnum: 95.31.152.0 - 95.31.255.255
  netname: BEELINE-BROADBAND
  descr: **Dynamic IP Pool** for Broadband Customers
  country: RU

- **213.169.149.3:**
  inetnum: 213.169.149.0 - 213.169.149.255
  netname: NETSHOP-ISP-LGNET-PA
  descr: 213.169.149.0-PA
  country: CY

What Does Spamhaus Say?

- **95.31.192.232 is listed on the Spamhaus SBL:**
  SBL223184 [**Yambo Financials**] ← YF dates to 2004!
  SBL227367 [Spammer hosting (escalation), /23]
  SBL243537 [ROKSO spammer hosting (escalation), /17]
  - http://www.spamhaus.org/rokso/spammer/SPM880/**yambo-financials**
  - https://www.spamhaus.org/rokso/sbl_**current**/SPM880/yambo-financials
  - https://www.spamhaus.org/rokso/sbl_**archived**/SPM880/yambo-financials

  See also:
  http://spamtrackers.eu/wiki/index.php/**Yambo_Financials**

- **213.169.149.3 is simply listed on the Spamhaus SBL** as "Spammer hosting"

VI. Applying Passive DNS to Our Example

Checking 95.31.192.232 in Passive DNS

- $ dnsdb_query.py -l 1000000 –i 95.31.192.232 > temp.txt
- $ wc -l temp.txt

  283624 temp.txt ← fully qualified domain names (more than 283K records is arguably too many to easily work with)

- $ more temp.txt

  [...]

  may.ajpxypim.be. IN A 95.31.192.232 ← wildcarding...

  arch.ajpxypim.be. IN A 95.31.192.232

  boon.ajpxypim.be. IN A 95.31.192.232

  coat.ajpxypim.be. IN A 95.31.192.232

  goad.ajpxypim.be. IN A 95.31.192.232

  knot.ajpxypim.be. IN A 95.31.192.232

  [etc]

Wildcarded Domain Names

- Wildcarding may be done by some opponents in an effort to frustrate attempts at automatically counting fully qualified domain names by some anti-spam mechanisms.

- Wildcarding may also have been done in an effort to frustrate use of passive DNS methods (some passive DNS sites limit the number of records returned to a relatively small number of records: if there are 100,000 wildcarded domains in an IP, but you can only see a maximum of 10,000 of them, well, that means passive DNS results may be incomplete. (Limits vary from passive DNS provider to passive DNS provider)

- Other times wildcarding may be used to track per-user "message opens," or in an attempt to locate instrumented sites

Can We De-Wildcard the FQDNs? Of Course...

- $ awk '{print $1}' < temp.txt > temp2.txt ← get just the names
- $ vi temp2.txt (decouple)
  :1,$s/\(^.*\.\)\(.*\)\(\.com\.$\)/\2\3/ ← de-wildcard dot coms
  :1,$s/\(^.*\.\)\(.*\)\(\.ru\.$\)/\2\3/ [etc]
  :1,$s/\(^.*\.\)\(.*\)\(\.co\.uk\.$\)/\2\3/ (condense)
  [etc]
  :wq
- $ sort temp2.txt | uniq > temp3.txt ← de-duplicate
- $ wc -l temp3.txt
  1251 temp3.txt ← unique 2nd level domain names
- $ more temp3.txt
  1st-drugstore.com.
  aawrivdj.com.
  [etc]

Unique 2nd Level Domains on 95.31.192.232

- **We found 1,251 unique 2nd level domains, including:**
  beingdrugstore.ru.
  bestcanadiancompany.ru.
  bestdealpills.ru.
  bestdrugsmarket.ru.
  bestdrugstore.ru.
  besthealingelement.ru.
  bestherbsquality.ru.
  bestmediafirst.ru.
  bestmedicalbargain.eu.
  bestmedicaloutlet.ru.
  bestmedicamentgroup.ru.
  drugcutpillsdrugs.ru.
  druggenericspharm.ru.
  drugherbalpharmacy.ru.
  drugprescriptionlevitra.ru.
  drugsdrugstorepills.ru.
  drugsherbalpill.ru.
  drugstoredrugsrx.ru.
  drugstoredrugstorehealth.ru.
  drugstorehealthtabs.ru.
  drugstorehomerxmeds.ru.
  drugstoremedsped.ru.
  [etc]
  Still a lot of domains, and this is just on *ONE* IP!

An Aside/Obvious Caution

- Deduping wildcarded domains is not w/o risk
- foo.example.com may go to one IP address
- bar.example.com may go somewhere else
- example.com (raw 2nd level domain) may go some third place
- That's okay, however: at the same time you're condensing domains in order to dump wildcards, you can use passive DNS to also retrieve ALL the IPs that *.example.com resolved to ☺

Checking 213.169.149.3 in Passive DNS?

- 213.169.149.3 was another IP our original domains resolved to...
- $ dnsdb_query.py -l 1000000 –i 213.169.149.3 > tempb.txt
- $ wc -l tempb.txt
  16141 tempb.txt
- $ awk '{print $1}' < tempb.txt > tempb2.txt
- $ vi temp2.txt
  :1,$s/\(^.*\.\)\(.*\)\(\.com\.$\)/\2\3/
  :1,$s/\(^.*\.\)\(.*\)\(\.ru\.$\)/\2\3/
  :1,$s/\(^.*\.\)\(.*\)\(\.co\.uk\.$\)/\2\3/
  [etc]
  :wq
- $ sort tempb2.txt | uniq > tempb3.txt
- $ wc -l tempb3.txt
  97 temp3.txt

Was It A "Waste" To Check That Second IP? No

- There were 1,251 domains in ns IP 1, and 97 in ns IP 2
- Combining the two and uniq'ing the combined set, we saw 1,259 total domains (this implies that we found 8 new unique domains by checking that 2nd name server):

fastdrugcompany.ru.	← MyCanadianPharmacy brand
globalwelnessshop.ru.	← Canadian Health&Care Mall
goodcarecompany.ru.	← Canadian Health&Care Mall
goodmedicinalshop.ru.	← MyCanadianPharmacy brand
magiccuringsupply.ru.	← Canadian Family Pharmacy
remedialsafeexchange.ru.	← Canadian Health&Care Mall
robinettewillamina.ru.	← [no logo]
wilhelminacal.ru.	← [no logo]

- We now know that this infrastructure isn't just in use by "Canadian Health&Care Mall"

Domain to IPs: Passively "Resolving" IPs

- When we resolved our original six domains, we only looked at what they *currently* resolved to (we used dig)...

- We could just as readily have looked at what they currently *and previously* resolved to in passive DNS...

- We can also check not just for one specific FQDN, we can actually look at ALL the wildcarded domains associated with a single level domain name...

For Example, What Do We See If We Check *.hwqdjzbw.eu? It's Been On 84 Different IPs...

- $ dnsdb_query.py -r *.**hwqdjzbw.eu**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp1.txt
- $ cat temp1.txt
  1.202.90.114
  103.16.198.47
  103.224.23.181
  103.238.216.219
  103.249.84.123
  103.31.251.244
  109.162.16.153
  109.251.176.16
  110.74.129.145
  111.184.108.234
  [etc]

We Can Find The IPs Used By Variants of Our Other Five Original Domains, Too..

- $ dnsdb_query.py -r *.**mymedicinalstore.in**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp2.txt
- $ dnsdb_query.py -r *.**besthealingelement.ru**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp3.txt
- $ dnsdb_query.py -r *.**purepharmacygrouponline.ru**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp4.txt
- $ dnsdb_query.py -r *.**homedrugquality.com**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp5.txt
- $ dnsdb_query.py -r *.**onlinecuringbargain.in**/A | grep -v ";;" | sed '/^$/d' | awk '{print $4}' | sort | uniq > temp6.txt
- $ cat temp1.txt temp2.txt temp3.txt temp4.txt temp5.txt temp6.txt | sort |uniq > temp-all.txt
- $ wc -l temp-all.txt

  298 temp-all.txt ← Total IPs tied to six original domains

Now Let's Find the Domains That Passive DNS Knows About Associated With Those 298 IP Addresses...

```
$ dnsdb_query.py -l 1000000 -i 1.202.90.114 > temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.16.198.47 >> temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.224.23.181 >> temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.237.33.165 >> temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.238.216.219 >> temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.247.211.241 >> temp-all-out.txt
$ dnsdb_query.py -l 1000000 -i 103.249.84.123 >> temp-all-out.txt
[etc]
$ wc -l temp-all-out.txt
1568595 temp-all-out.txt
```

How Many Records Per IP?

- 283780 95.31.192.232 [one of our original IPs]
  237615 178.156.236.5 [LEX Media Concepts SRL, RO]
  218151 213.186.33.3 [cluster015.ovh.net, FR]
  90111 213.155.21.33 [33.21.155.213.hosting.ua]
  74092 193.105.154.202 [CityNet Line, LV]
  69444 193.105.154.203 [ditto]
  52106 103.16.198.47 [PT. Jupiter Jala Arta, ID]
  46358 85.17.138.159 [Leaseweb, NL]
  25458 85.95.236.38 [Inetmar Internet, TR]
  25100 212.66.44.121 [Ints Internet Host, Dialup_1, UA]
  19480 91.198.137.39 [Adrian Andreas Zajonz/Host1, DE]
  17989 61.178.118.4 [Chinanet Gansu province, CN]
  16862 178.63.48.181 [Hetzner Datacenter 12, DE]
  [etc]

How Many FQDNs Per Domain?

- 89954 googlesydnication.com.
  12092 radioshackstores.in.
  10020 doctordi.com.ua.
  7700 pureremedialassist.com.
  7041 magicpharmacyassist.be.
  6971 yourtabstrade.eu.
  6726 besthealingelement.ru.
  6588 genericrxassist.be.
  6337 magicaidmart.eu.
  5923 herbalsecuremart.eu.
  5892 hwqdjzbw.eu.
  5741 familyhealthquality.be.
  5487 healingsmartgroup.be.
  5440 canadiandrugreward.eu.
  5384 hotdrugsstore.in.
  5360 bestmedicalbargain.eu.
  5321 drrespharmacy.in.ua.
  4958 wxbereho.in.
  4912 naturaldrugsmall.be.
  4857 vzbhtapi.eu.
  4825 dingdoctor.ru.
  4768 doctorla.com.ua.
  4757 organicsmartbargain.eu.
  4714 perfectaideshop.be.
  4243 pogjlbwb.eu.
  [etc]

How Many 2nd Level Domains?

- $ awk '{print $1}' < temp-all-out.txt | sort | uniq > temp-all-out2.txt
- $ vi temp-all-out2.txt
  :1,$s/\(^.*\.\)\(.*\)\(\.com\.$\)/\2\3/
  :1,$s/\(^.*\.\)\(.*\)\(\.ru\.$\)/\2\3/
  :1,$s/\(^.*\.\)\(.*\)\(\.co\.uk\.$\)/\2\3/
  [etc]
  :wq
- $ sort temp-all-out2.txt | uniq > temp-all-out3.txt
- $ wc -l temp-all-out-3.txt
  233348 temp-all-out-3.txt
- Are all those domains prima-facie tied to subject matter of interest? No. Some appear to be unrelated (but it can be hard to tell w/o visiting the domains)

An Explosion of Complexity... and False Positives?

- Being able to go from one IP address to many additional IPs and domain names is huge: that's a **lot** of additional leads for an investigator to potentially chase!
- However, as we go further and further from our original "clues," the risk goes up that we'll find stuff that may not be obviously related
- You may want to investigate whois data for non-obvious domains, at least if it isn't privacy/proxy protected

- Or we could begin by just pulling the "usual suspects"

The "Usual Suspects"

- $ cat suspects
  apoteket
  apotheke
  canadian
  cialis
  doctor
  drug
  generic
  [etc]

🡐 sorry folks ☹

- $ grep -f suspects temp-all-out3.txt > suspect-domains.txt
- $ wc -l suspect-domains.txt
  25915 suspect-domains.txt

Focusing on Leftovers

- Once you've identified domains that match "obvious suspects," save them, then remove them what's left over:

- Easy weedout script for this:

```
#!/usr/local/bin/perl
while (<>) {
        $weed{$_}=1;
}
while (<>) {
        print unless $weed{$_};
}
```

- $ weedout suspect-domains.txt < temp-all-out3.txt > leftovers.txt

Random Domains, Innocent-Looking Domains, Etc

- But what about the "random"-looking domains such as hwqdjzbw.eu ? Or even totally innocent looking domains nonetheless used for unsolicited email or other bad purposes?
- You may only be interested in domains that currently resolve; that's one easy way to weed out at least some additional domain names (but beware of deceptive name severs lying to you, and of course, if you get to obvious, you may be "noticed" by the bad guys, and they may bolt entirely)
- Another option is to check the domains that you've found against domain name block lists such as SURBL or the Spamhaus DBL, only keeping those that are listed by SURBL or Spamhaus
- However, even if you do nothing more, you've still gone from half a dozen domains to 25,915 presumptively evil domains

VII. The Special Case of Name Server Names

Name Servers Used By The Observed Domains

- hwqdjzbw.eu. → ns1.vnjreuhl.be.
  hwqdjzbw.eu. → ns2.qjajcjzv.ru.
- mymedicinalstore.in. → ns{1,2}.mymedicinalstore.in.
- besthealingelement.ru. → ns{1,2}.besthealingelement.ru.
- purepharmacygrouponline.ru. → ns{1,2}.purepharmacygrouponline.ru.
- homedrugquality.com. → ns{1,2}.homedrugquality.com.
- onlinecuringbargain.in: → ns{1,2}.onlinecuringbargain.in.

- "Name-servers-that-match-the-base-domain-name" say "please check the IP addresses underlying those name servers..." [miscreants normally do not go to the trouble of using unique IPs for each unique name server, they multiplex]

So We Begin By Looking For Commonalities

- Just as we did with the original half dozen domain names, we begin by looking for commonalities in the name server IPs

Name Servers Used By The Observed Domains

- hwqdjzbw.eu. ➔ ns1.vnjreuhl.be. ➔ **45.64.105.144**
  hwqdjzbw.eu. ➔ ns2.qjajcjzv.ru. ➔ **103.249.84.123**
- mymedicinalstore.in. ➔ ns1.mymedicinalstore.in. ➔ **45.64.105.144**
  mymedicinalstore.in. ➔ ns2.mymedicinalstore.in. ➔ **103.249.84.123**
- besthealingelement.ru. ➔ ns1.besthealingelement.ru. ➔ **167.88.8.9**
  besthealingelement.ru. ➔ ns2.besthealingelement.ru. ➔ **103.249.84.123**

- 45.64.105.144 IS listed by Spamhaus
- 103.249.84.123 IS listed by Spamhaus [as Yambo Financials]
- 167.88.8.9 IS listed by Spamhaus

Are There Other Name Server FQDNs on the 45.64.105.144 NS Address? Yes – 86 of Them

- Name servers on that IP include:

  ns1.eanjjpgn.ru. IN A 45.64.105.144
  ns1.globalremedytrade.ru. IN A 45.64.105.144
  ns1.healingpharmgroup.ru. IN A 45.64.105.144
  ns1.hwqdjzbw.eu. IN A 45.64.105.144
  ns1.ixoowbuh.com. IN A 45.64.105.144
  ns1.jbccdqsf.com. IN A 45.64.105.144
  ns1.joystiqdeals.ru. IN A 45.64.105.144
  ns1.kbsvyvbp.com. IN A 45.64.105.144
  ns1.kbsvyvbp.com. IN A 45.64.105.144
  [etc]

- Each of *those* name servers can in turn be sent through passive DNS to see what they resolve to...

The 86 Name Servers Expand to 1,608 Unique NS Records Due to Multiple "A" Recs Per FQDN

- ns1.eanjjpgn.ru. IN A 167.88.8.9
  ns1.eanjjpgn.ru. IN A 188.120.245.240
  ns1.eanjjpgn.ru. IN A 45.64.105.144
  ns1.globalremedytrade.ru. IN A 103.16.198.47
  ns1.globalremedytrade.ru. IN A 45.64.105.144
  ns1.healingpharmgroup.ru. IN A 103.16.198.47
  ns1.healingpharmgroup.ru. IN A 103.224.23.181
  ns1.healingpharmgroup.ru. IN A 167.88.8.9
  ns1.healingpharmgroup.ru. IN A 178.239.176.235
  ns1.healingpharmgroup.ru. IN A 190.123.45.111
  ns1.healingpharmgroup.ru. IN A 212.57.38.19
  ns1.healingpharmgroup.ru. IN A 45.64.105.144
  ns1.hwqdjzbw.eu. IN A 1.202.90.114
  [etc]

Let's Just Look At the IP's From Those: 700 Unique IPs

- 1.202.90.114
  103.16.198.47
  103.224.23.181
  103.224.23.7
  103.23.241.212
  103.237.33.165
  103.238.216.219
  103.241.150.190
  103.247.211.241
  103.249.84.123
  103.253.114.124
  103.8.27.30
  107.20.157.100
  [etc]

Checking Passive DNS for the Domains On Those IPs: 4,429,062 Unique Records...

- ns1.dfnfoskt.be. IN A 1.202.90.114
  ns2.dfnfoskt.be. IN A 1.202.90.114
  ns1.vrswovjt.be. IN A 1.202.90.114
  ns2.vrswovjt.be. IN A 1.202.90.114
  ns1.naturaldrugsmall.be. IN A 1.202.90.114
  ns2.naturaldrugsmall.be. IN A 1.202.90.114
  coarse.naturaldrugsmall.be. IN A 1.202.90.114
  forger.naturaldrugsmall.be. IN A 1.202.90.114
  gurgle.naturaldrugsmall.be. IN A 1.202.90.114
  intend.naturaldrugsmall.be. IN A 1.202.90.114
  [etc]
- If we simplify those, how many unique 2nd level domains?
  344,048

Distribution of FQDN Records Per IP Address

- 700,808 124.248.205.53
  586,064 124.109.1.160
  281,193 103.253.114.124
  237,615 178.156.236.5
  217,659 213.186.33.3
  198,500 91.240.165.35
  186,954 94.62.13.10
  149,714 122.0.69.152
  93,144 116.255.159.117
  90,438 119.84.74.32
  90,108 213.155.21.33
  74,089 193.105.154.202
  69,444 193.105.154.203
  52,106 103.16.198.47
  46,358 85.17.138.159
  40,620 61.133.234.105
  39,900 217.23.14.174
  37,221 87.107.121.213
  36,631 65.254.248.136
  32,163 212.83.186.116
  31,344 31.184.224.133
  31,318 85.25.253.145
  30,754 91.109.18.86
  27,547 207.182.155.101
  25,458 85.95.236.38
  21,083 211.20.209.245
  19,987 198.245.62.48
  [etc]

What's Known About The Top IPs?

- 124.248.205.53 ("SunnyVision Limited," HK)
  124.109.1.160 ("BangkokVPS Cloud Server," TH)
  103.253.114.124 ("PT Media Andalan Nusa," ID)
  178.156.236.5 ("LEX Media Concepts SRL," RO)
  213.186.33.3 ("OVH," FR)
  91.240.165.35 ("Il Cupola Verboom B.V.," NL)
  94.62.13.10 ("Vodafone Telecel," PT)
  122.0.69.152 ("HuaBei Oil Communication Co.," CN)
  116.255.159.117 ("ZhengZhou Giant Computer Network," CN)
  119.84.74.32 ("Chinanet Chongqing Province," CN)
  213.155.21.33 ("ALMGHARII7 - Mohammed Almghari," UA)
  193.105.154.202 ("CityNet Line/ARS Tolerantia," LV)
  193.105.154.203 ("CityNet Line/ARS Tolerantia," LV)
  103.16.198.47 ("PT Jupiter Jala Arta," ID)

How Many Records Per 2nd Level Domain Name? (Domain Wildcarding and/or Multiple IP Effects)

- 89954 googlesydnication.com.
  84418 googlsyndications.com.
  10020 doctordi.com.ua.
  8896 medichump.ru.
  6992 radioshackstores.in.
  6340 doctorknob.ru.
  6339 keysmedic.ru.
  6332 doctormelt.ru.
  6330 ru.com.
  6236 doctordell.ru.
  6223 doctorowed.ru.
  6065 doctorclog.ru.
  5796 flagshere.com.
  5758 bitdrugs.ru.
  5635 comprisingmeds.pl.
  5626 doctortyro.ru.
  5597 pillspeer.ru.
  5463 doctortorn.ru.
  5381 doctorreek.ru.
  5374 doctorgobs.ru.
  5368 tabletsmedshealth.ru.
  5247 wxbereho.in.
  5223 medicpace.ru.
  4973 familyhealthquality.be.
  4900 awaitaltogether.pl.ua.
  4864 dingdoctor.ru.
  [etc.]

We Could Go A LOT Deeper On Name Servers

- We only looked at name servers associated with our initial six domains and the IPs we saw for them:
  Initial six domains ➔
  Find name servers used by those six domains ➔
  Map name servers to 3 IP addresses ➔
  86 total name servers on those 3 IPs ➔
  1,608 unique NS/IP records ➔
  700 unique IPs ➔
  4,429,062 unique FQDN+IP records
  344,048 unique 2nd level domains

- That was for just 86 name servers.
- I've got a list of 39,953 related name servers we could check...

Bottom Line...

- You can iterate through passive DNS finding new potentially related domain names and IP addresses and name servers
- You can end up finding a LOT of domains and IPs and name servers this way
- **But when should you stop?**

VIII. Oh Please, Let It Be Quitting Time...

Once you learn to quit,
it becomes a habit.

Vince Lombardi

Avoiding Analyst Obsessiveness

- An analyst could potentially become obsessed looping through passive DNS results

- We hope, therefore, to offer some suggestions for eventually terminating one's analysis and avoiding a "descent into madness and obsession."

- Note that these suggestions may result in you prematurely terminating your explorations, so feel free to use your best judgment and disregard these suggestions. (I certainly don't want to discourage initiative!)

Pick A Target *Before* You Start

- This is like deciding how much you're willing to lose gambling, long before you ever sit down at a poker table. For example, maybe you're fine spending a hundred bucks in a casino for an evening's entertainment; similarly, maybe you'd be happy if you could identify even 25,000 related domains.
- Having an initial target is one of the most sane approaches you might consider, particularly if you're looking at requesting domain take downs or you're building a case for prosecution.
- By picking a relatively small number you can be selective about the domains you pick, ensure you get particularly fresh examples, or do particularly good work checking and writing them up (but don't become such a perfectionist that you never get beyond working a couple hundred domains!)

Plan to Work on Multiple Discrete Tranches

- A variant of the pick-a-fixed target set size is to consider working on a series of batches, or tranches, of domains.

- For example, you might do weekly or monthly "batches," recognizing that most (large) spam problems will require multiple months-worth of sustained effort.

Six Degrees of Kevin Bacon*

- If you've used passive DNS to go:

  names → numbers
  numbers → names

  through **six iterations,** you should stop. Your chances of hitting stuff that is only tangentially related has become quite high.

- The more you want to feedback one cycle into another, the better you need to be about whitelisting the unlistable (for example: Google, Apple, Microsoft, CNN, Amazon, etc.)

* http://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon

3 Million Unique 2nd Level Domain Names

- In 2015, there are just over 288 million domain names worldwide.*

- If your analysis has resulted in you finding more than 3 million unique 2nd level domain names, you're working with more than 1% of all global domain names.

- That's enough for any purpose I can think of. You should stop. Work on improving what you've already got.

---

* http://www.verisigninc.com/en_US/innovation/dnib/index.xhtml

Stale Data

- If you're finding lots of domains that no longer resolve, you may want to consider time fencing your queries.

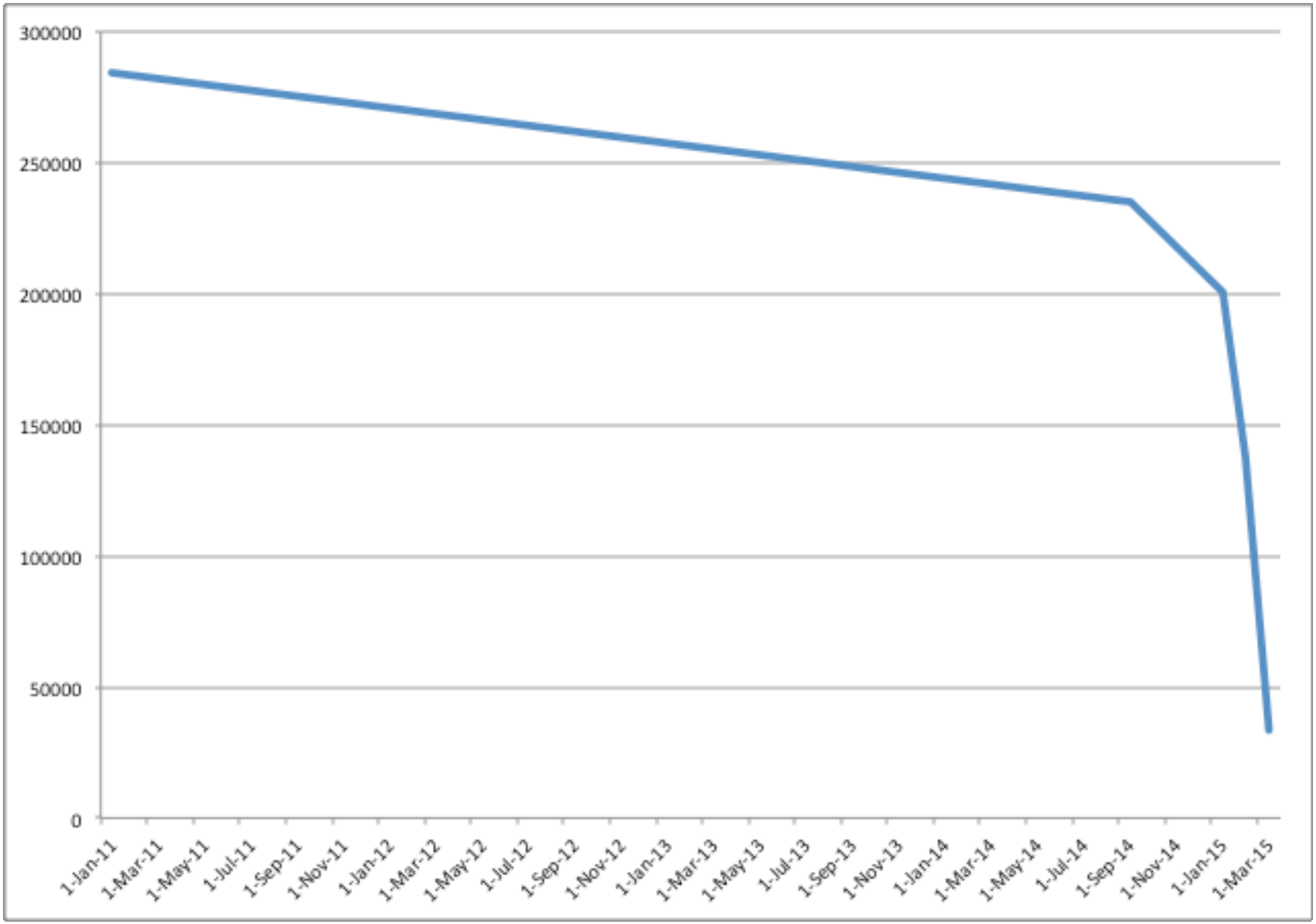
- Data from three years ago may be of academic interest, but if the domain no longer exists, and the IP addresses are now serving someone else's pages, it probably isn't directly actionable.

- Think about a six month window, or even a three month window, or even a one month window...

Limiting Records Returned For 95.31.192.232

- $ dnsdb_query.py -i 95.31.192.232 | wc -l
  284220
- $ dnsdb_query.py -l 1000000 **--after 2014-09-01** -i 95.31.192.232 | wc -l
  235331
- $ dnsdb_query.py -l 1000000 **--after 2015-01-01** -i 95.31.192.232 | wc -l
  200386
- $ dnsdb_query.py -l 1000000 **--after 2015-02-01** -i 95.31.192.232 | wc -l
  138060
- $ dnsdb_query.py -l 1000000 **--after 2015-03-01** -i 95.31.192.232 | wc -l
  33548

Time Fencing 95.31.192.232



Comparison of Found Results Against Known Listings

- You could also use known data points associated with an identified entity (such as Spamhaus' listing of currently known Yambo Financial SBLs) and use that as a basis for comparing what you've already found via passive DNS to what Spamhaus knows about this outfit.
- If, for example, you were to check passive DNS for all the SBL listed Yambo Financial IPs and CIDR blocks, and no additional domains or IPs were to turn up, you might feel fairly comfortable that your coverage, even if potentially/theoretically incomplete, is practically quite all-inclusive and impressive.
- Of course, you might also consider "cheating" and using those listing as initial "seeds" for passive DNS research (although at that point they'd obviously lose their evaluative value!)

High-Value (Strategic) Derived Insights

- Others may choose to focus on particularly valuable insights, stopping when they've found something that's particularly strategic.
- For example, if there's a single payment processor site that's used by thousands of different domains, successfully going after that single site may be more valuable than going for bulk take downs.
- Similarly, if most throw away sites are actually just a reverse proxy, it may be worthwhile attempting to identify the site that is actually serving as the back end for all the front end reverse proxies.
- It can be hard to take this sort of long-term strategic view.

Thanks for the Chance to Talk Today!

- We hope you're now inspired to try passive DNS methods in your own cyber investigations! May you find yourself awash in new leads!

- Are there any questions?