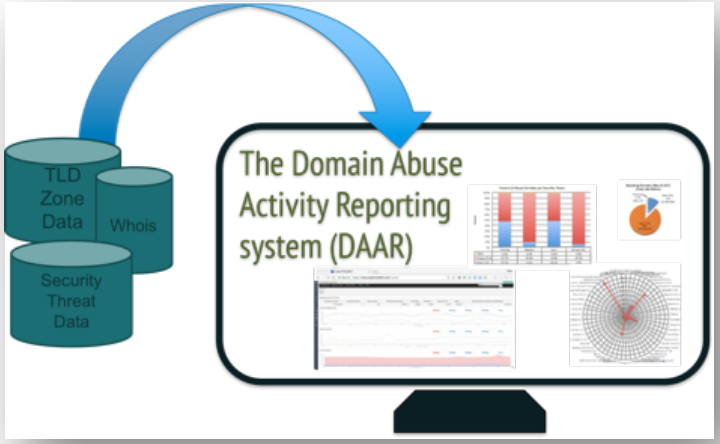
The Domain Abuse Activity Reporting System (DAAR)



Dave Piscitello

APWG EU
October 2017

The Domain Abuse Activity Reporting system

What is the Domain Abuse Activity Reporting system?

- A system for reporting on domain name registration and abuse data across TLD registries and registrars

How does DAAR differ from other reporting systems?

- Studies all gTLD registries and registrars for which we can collect zone and registration data
- Employs a large set of reputation feeds (e.g., blocklists)
- Accommodates historical studies
- Studies multiple threats: phishing, botnet, malware, spam
- Takes a scientific approach: transparent, reproducible

DAAR & the Open Data Initiative

- Goal of Open Data Initiative is to facilitate access to data that ICANN organization or community creates or curates
- DAAR system uses data from public, open, and commercial sources
  - DNS zone data
  - WHOIS data
  - Open source or commercial reputation blocklist (RBL) data
    - Certain data feeds require a license or subscription
- In cases where licensing permits, DAAR data or reports will be published and included in the Open Data Initiative

Project Goals

- DAAR data can be used to
  - Report on threat activity at TLD or registrar level
  - Study histories of security threats or domain registration activity
  - Help operators understand or consider how to manage their reputations, their anti-abuse programs, or terms of service
  - Study malicious registration behaviors
  - Assist operational security communities

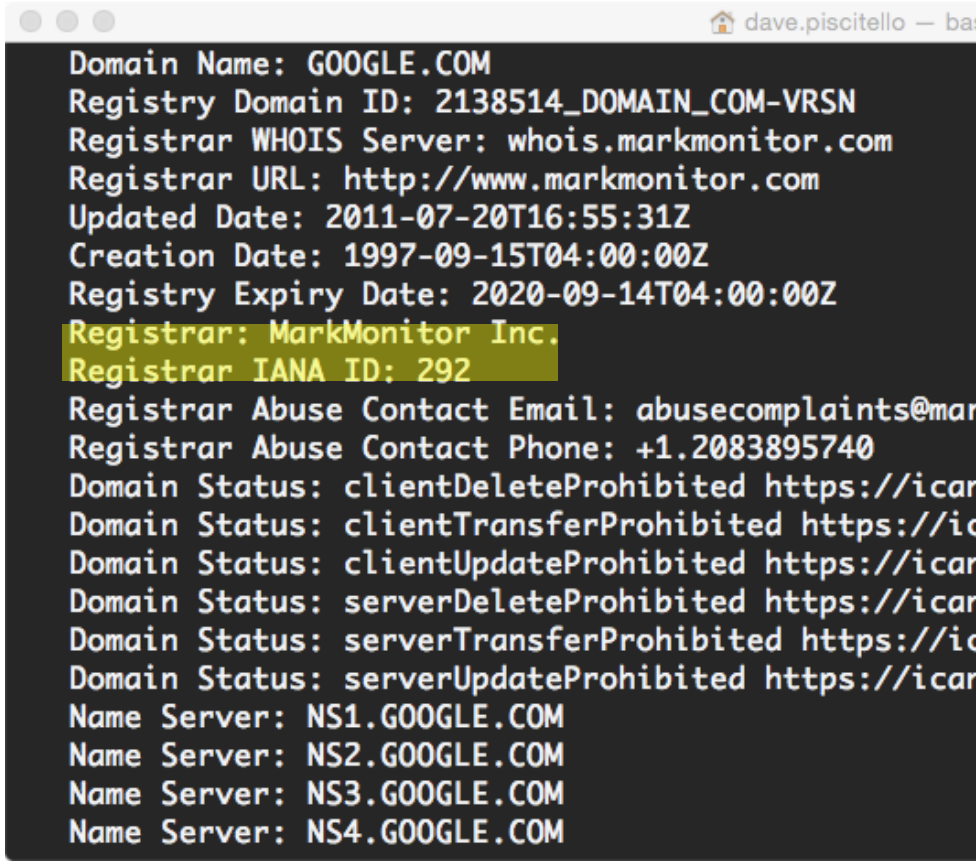
The purpose of DAAR is to provide data to support community, academic, or sponsored research and analysis for informed policy consideration

DAAR Uses TLD Zone Data

- Collects all gTLD zones for gTLD registry analytics
- DAAR uses publicly available methods to collect zone data
  - Centralized Zone Data Service, zone transfer)
- DAAR only uses domain names that appear(ed) in zones
- Currently, system collects zones from ~1240 gTLDs
  - Approximately 195 million domains

DAAR Uses Whois

- DAAR uses published registration data (Whois)
  - Uses only registration data necessary to associate resolving domain names in zone files with sponsoring registrars
- Reliable, accurate registrar reporting depends on Whois
  - Collecting registration records for millions of domains is a big challenge

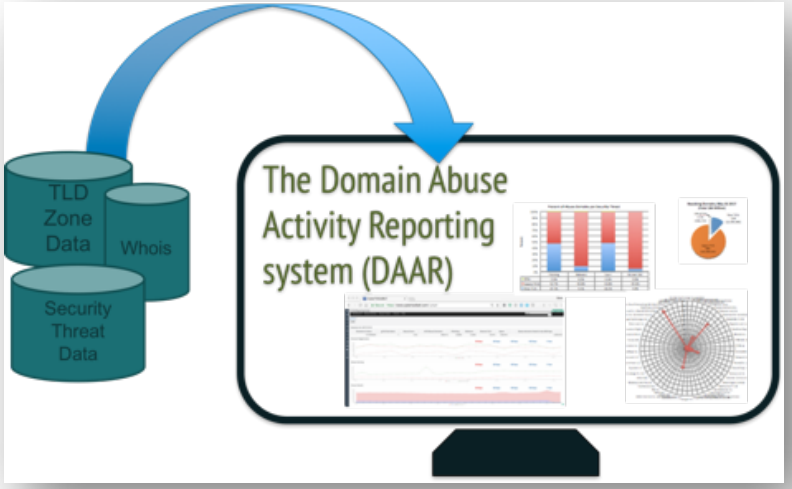


DAAR Uses Many Threat Data Sets

- DAAR counts “unique” abuse domains
  - A domain that appears on *any* RBL reporting to DAAR is included in the counts *once*
- DAAR uses multiple domain or URL abuse data sets to
  - Generate daily counts of domains associated with phishing, malware hosting, botnet C&C, and spam
  - Calculate daily total and cumulative abuse domains
  - Calculate newly added abuse domains (a monthly count), and cumulative abuse domains (365 day count)
  - Create histograms, charts, days in the life views

DAAR reflects how entities external to ICANN community see the domain ecosystem

Reputation Data: Identifying Threats



DAAR Is Not An Abuse List Service

- OCTO-SSR does *not* compose its own reputation blocklists
  - DAAR presents a composite of the data that external entities use to block threats
- DAAR collects the same abuse data that is reported to industry and Internet users
  - The abuse data that DAAR collects are used by commercial security systems that protect millions of users and billions of mailboxes daily
  - Academic and industry use and trust these data sets
  - Academic studies and industry use validate these data sets exhibit accuracy, global coverage, reliability and low false positive rates

DAAR Criteria for Reputation Data (RBLs)

- RBLs must provide threat classification that match our set of security threats
- Evidence that operational and security communities trust the RBL for accuracy, clarity of process
- RBLs have positive reputations in academic literature
- RBLs are broadly adopted across operational security community
  - Feeds are incorporated into commercial security systems
  - Used by network operators to protect users and devices
  - Used by email and messaging providers to protect users

Reputation Block Lists: Protecting Users Everywhere

- RBL use is nearly ubiquitous
- RBLs block more than unsolicited commercial email
- RBLs in Browsers
  - Google Chrome uses APWG, and Safe Browsing URL Data
- RBLs in the Cloud and Content-Serving Systems
  - Akamai uses SURBL, Symantec, ThreatSTOP, and custom RBLs
  - AWS WAF uses RBLs to block abuse or volumetric attacks
  - Google Safe Browsing blocks malicious URLs and AdWords fraud
- RBLs in Your Social Media Tools
  - Facebook composes and shares its ThreatExchange platform
- RBLs in the DNS
  - ISPs & private networks use Resource Policy Zones (RPZs) at resolvers.
  - Spamhaus and others provide RBLs in RPZ format

Reputation Block List Uses: Private Network Operators

- RBLs in commercial firewalls, UTM devices
  - Admin guides from Palo Alto Networks, Barracuda Networks, SonicWall, Check Point, Fortigate, Cisco IronPort, and WatchGuard
  - TitanHQ SpamTitan, Sophos UTM, andProofpoint also provide RBL-based filtering to protect users from visiting malicious URLs
  - External RBLs mentioned: Spamhaus, SURBL, SpamCop, Invaluement, abuse.ch, Open ORDBL, Spam and Open Relay Blocking System (SORBS), Squidblacklist.org,
- RBLs in enterprise mail/messaging systems
  - Spam solutions from GFI MailEssentials, SpamAssassin, and Vamsoft ORF include Spamhaus or SpamCop RBLs available for Microsoft Exchange
- RBLs and Third-Party Email Service Providers (ESPs)
  - Amazon Simple Email Service RBL or DNS block lists
  - Look at ESPMail Exchange (MX) and Sender Policy Framework (SPF) resource records

RBLs in Academia: One means of asserting RBL confidence

Partial list of academic studies and citations of RBLs that report to DAAR

Empirically Characterizing Domain Abuse and the Revenue Impact of Blacklisting

Blacklist Ecosystem Analysis: Spanning Jan 2012 to Jun 2014

Taster's Choice: A Comparative Analysis of Spam Feeds

Learning to Detect Malicious URLs

Understanding the Domain Registration Behavior of Spammers

The Statistical Analysis of DNS Abuse in gTLDs (SADAG) Report

Shades of grey: On the effectiveness of reputation-based blacklists

Click Trajectories: End-to-End Analysis of the Spam Value Chain

Current Reputation Data Sets

- SURBL lists (domains only)
- Spamhaus Domain Block List
- Anti-Phishing Working Group
- Malware Patrol (Composite list)
- Phishtank
- Ransomware Tracker
- Feodotracker

SpamAssassin: malware URLs list
Carbon Black Malicious Domains
Postfix MTA
Squid Web proxy blocklist
Symantec Email Security for SMTP
Symantec Web Security
Firekeeper
DansGuardian
ClamAV Virus blocklist
Mozilla Firefox Adblock
Smoothwall
MailWasher

Does DAAR Identify All Abuse?

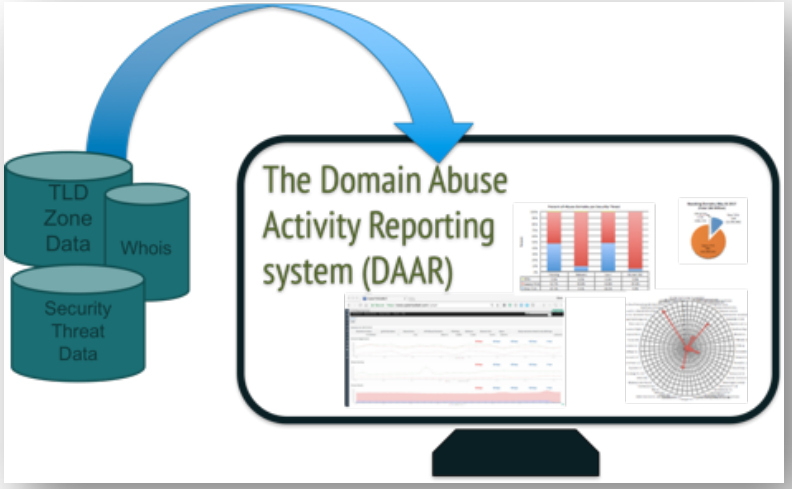
- No reputation provider can see all the abuse
  - Each is catching only some (what they see)
- Providers look for different types of abuse, use different methods or infrastructures
- Some lists are big and some are small.
  - The smaller the list, the less percent of overlap it might have with a larger list

Why Is DAAR Reporting Spam Domains?

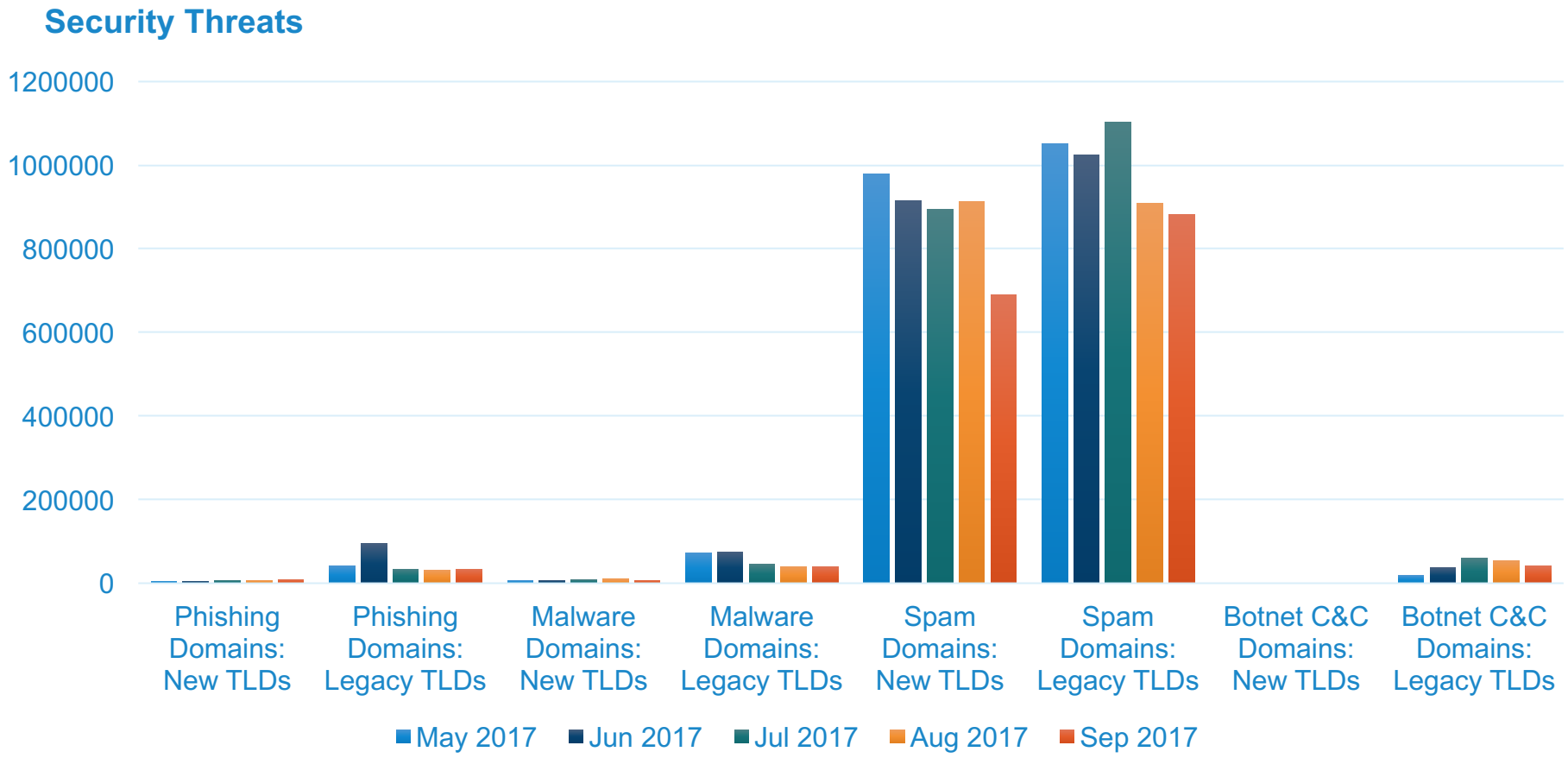
- The ICANN Governmental Advisory Committee (GAC) expressed interest in spam domains as a security threat in its Hyderabad correspondence to the ICANN Board of Directors… Why? Because
- Most spam are sent via illegal or duplicitous means (e.g., via botnets).
- Spam is no longer singularly associated with email
  - Link spam, spamdexing, tweet spam, messaging spam (text/SMS)
- Spam is a major means of delivery for other security threats
  - Spam has evolved to a (cloud) service: Avalanche, for example, provided domain registrations to customers
- DAAR mainly measures domain names found in the bodies of spam messages

MOST IMPORTANTLY, spam domain reputation influences how extensively or aggressively security or email administrators apply filtering

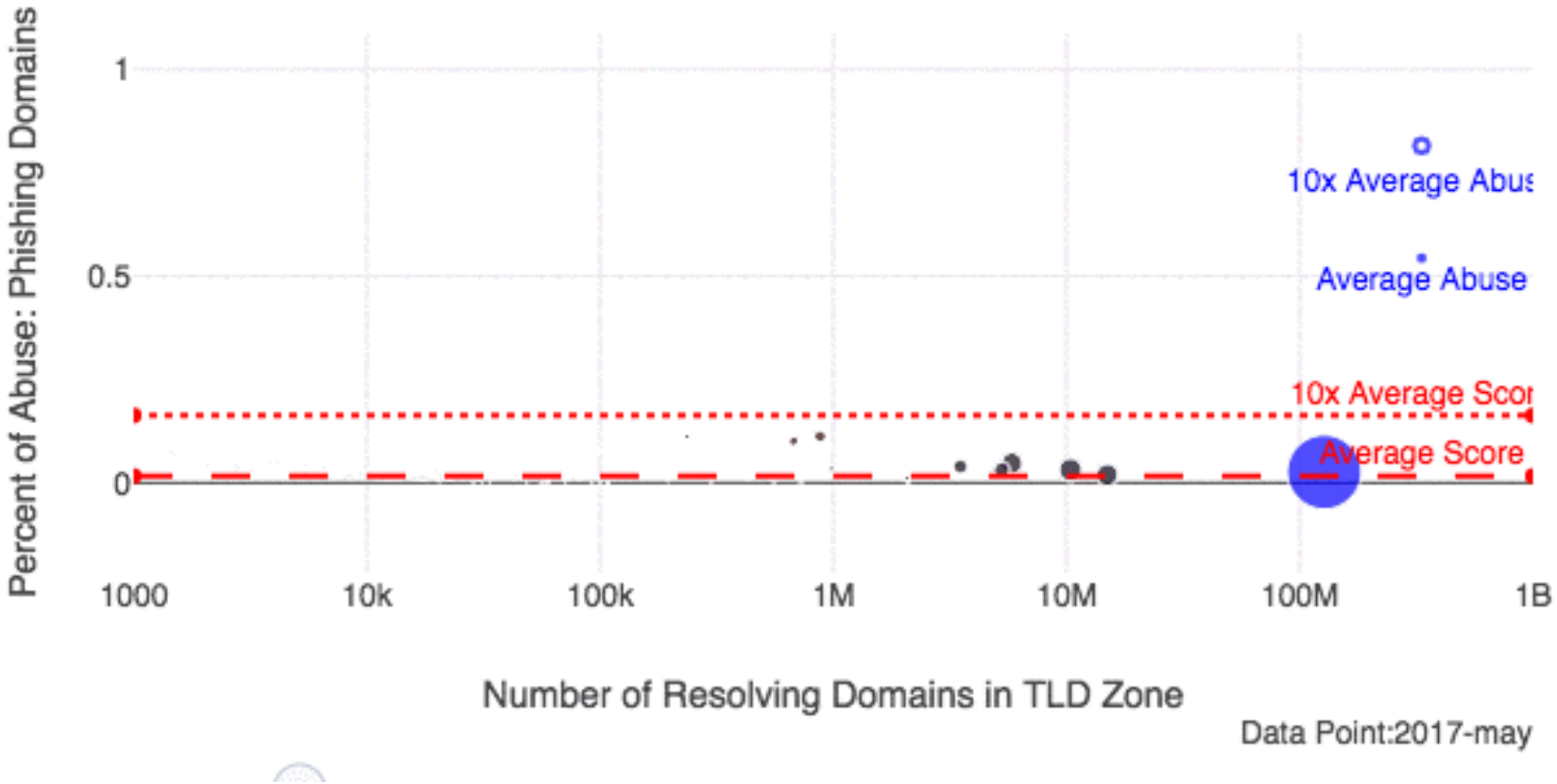
Visualizing DAAR Data



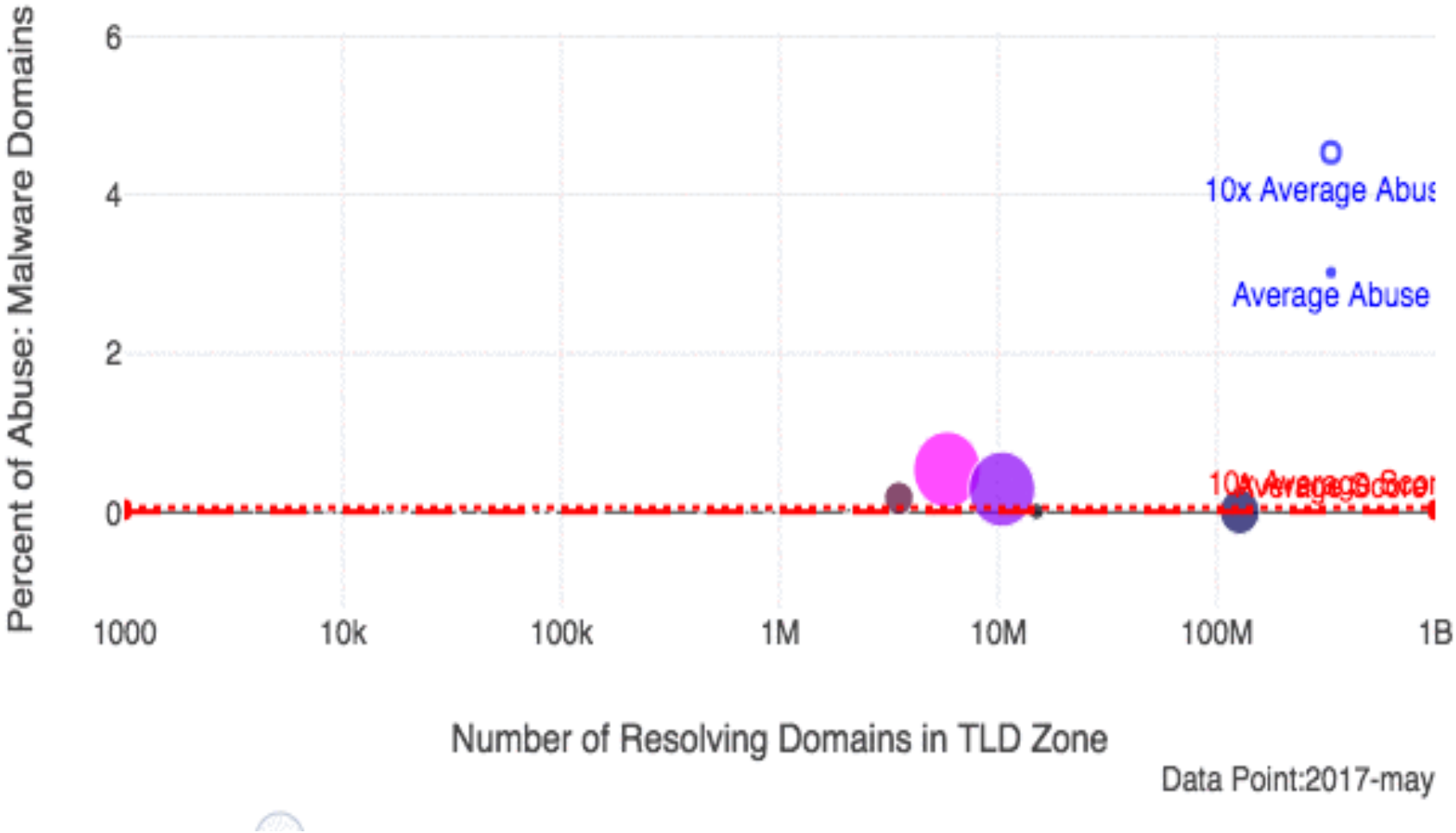
Data Set: All gTLDs having at least 1 reported abuse domain



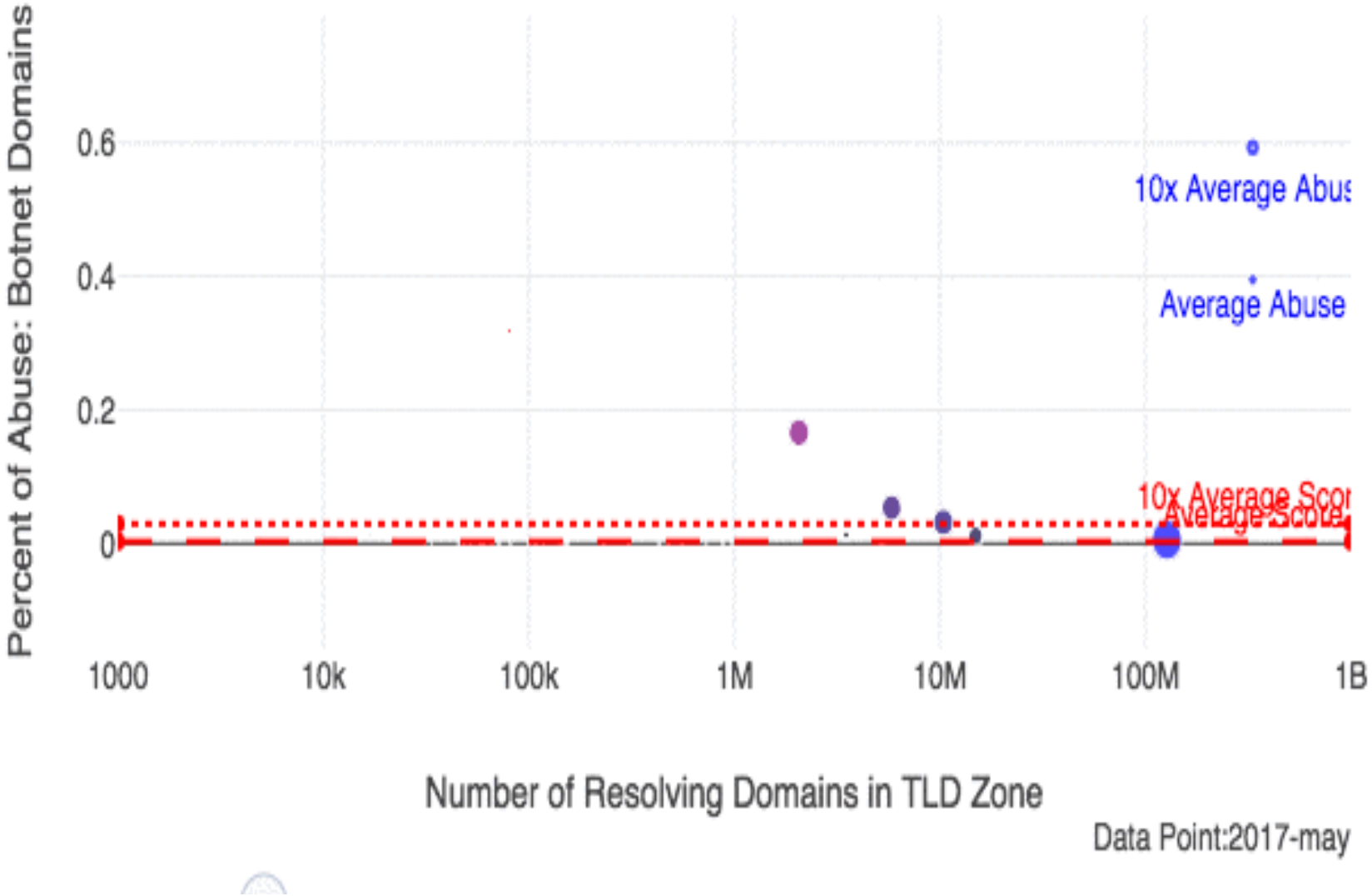
End of Month Snapshots: Phishing Domains Percent of Abuse



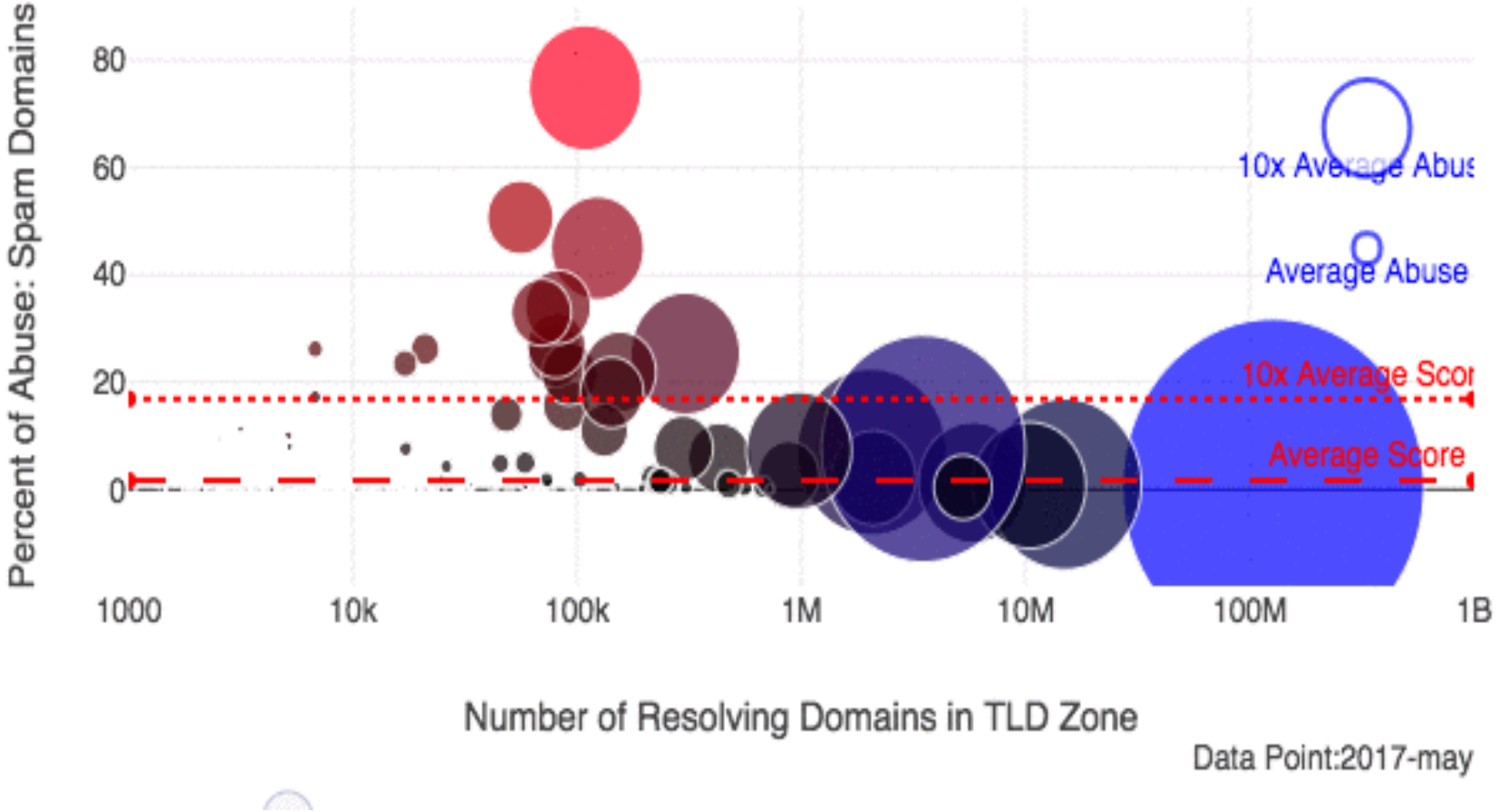
End of Month Snapshots: Malware Domains Percent of Abuse



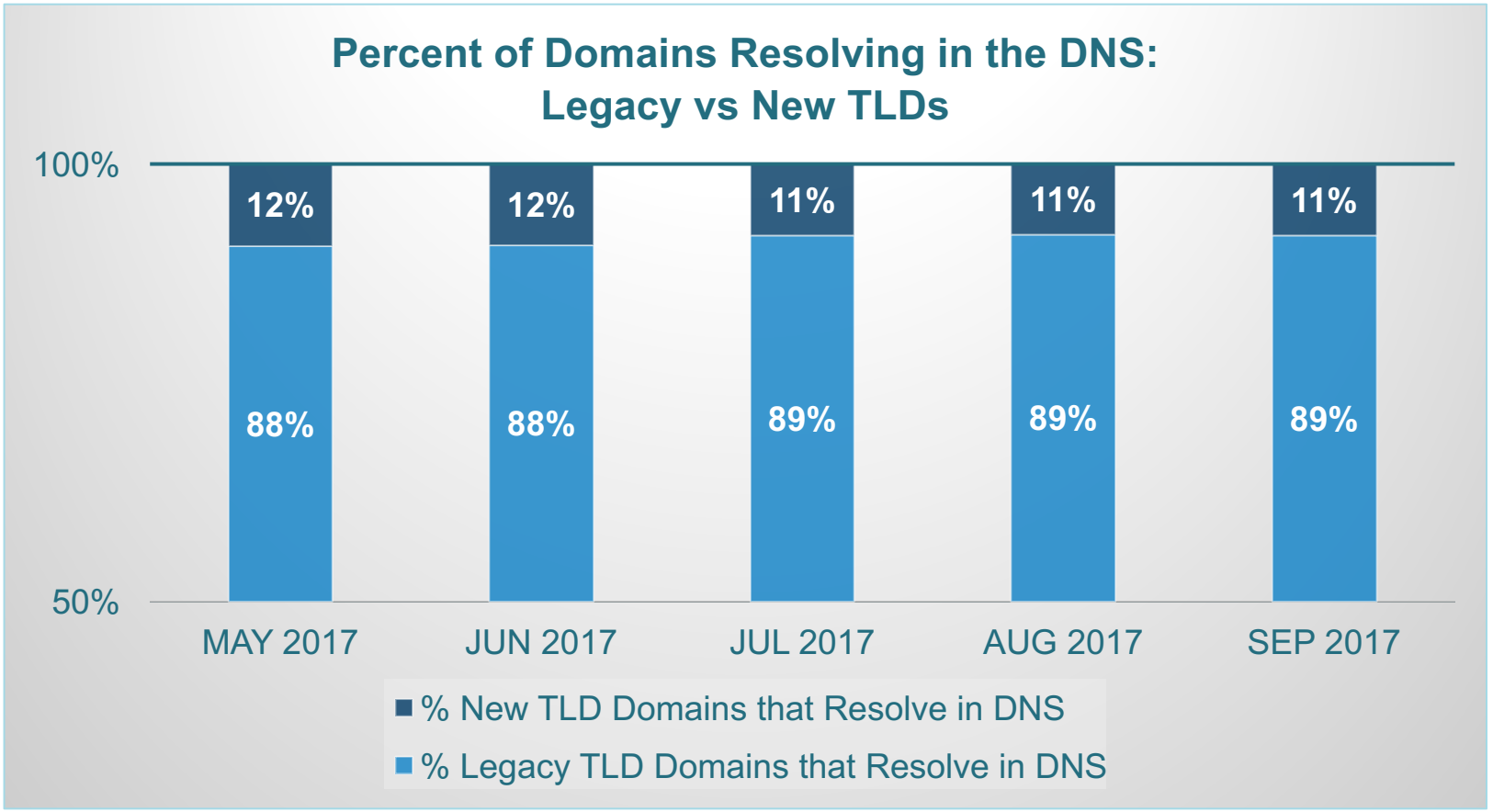
End of Month Snapshots: Botnet (C2) Domains Percent of Abuse



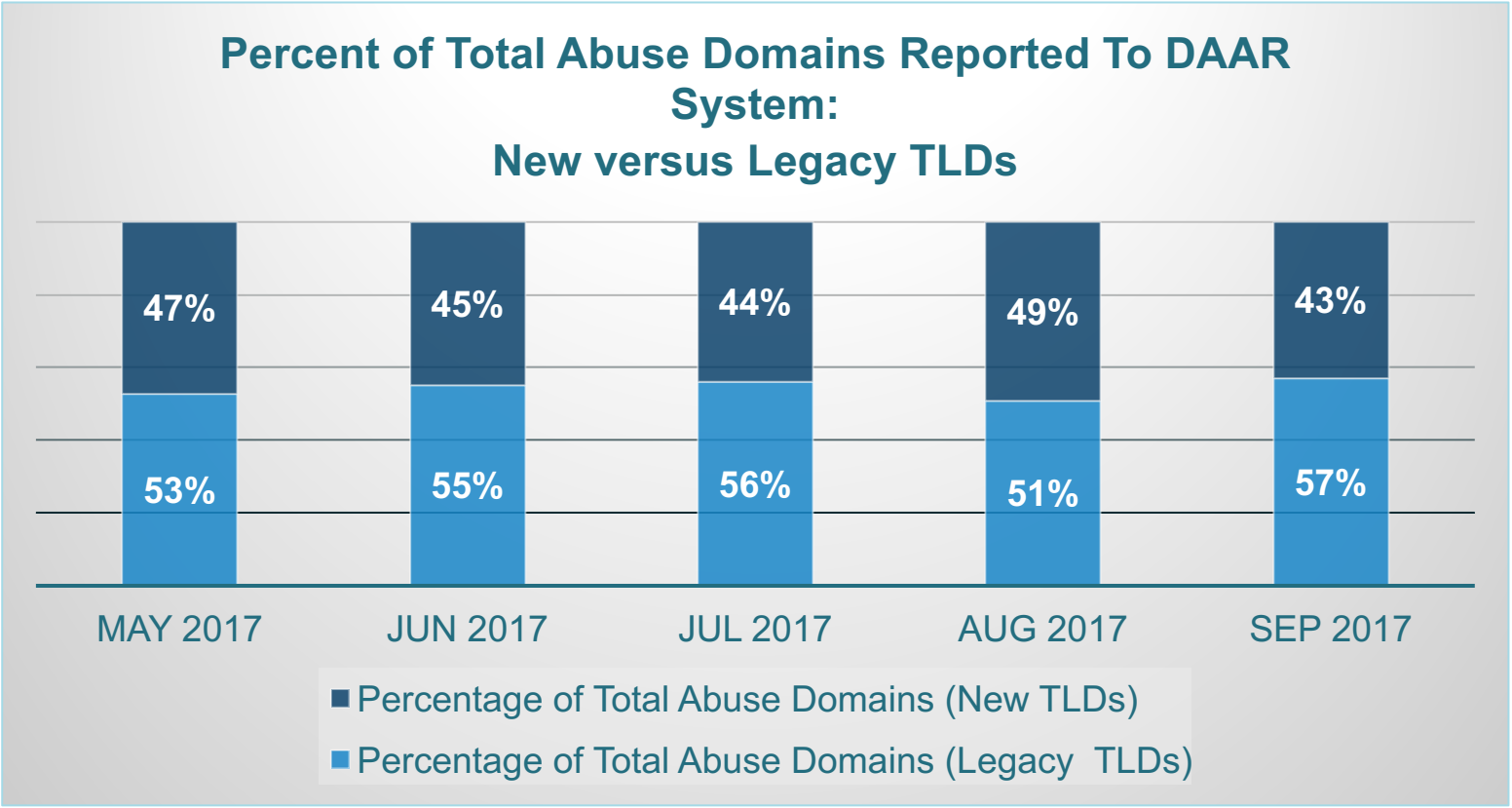
End of Month Snapshots: Spam Domains Percent of Abuse



Data Set: All gTLDs having at least 1 reported abuse domain



Data Set: All gTLDs having at least 1 reported abuse domain



Total Abuse Tells Only *PART* of the Story:
Let's drill down to Consider Concentration Or Distribution

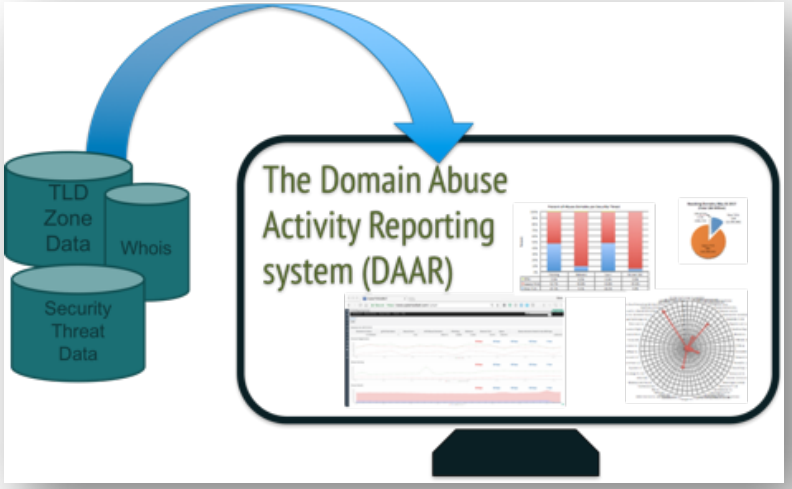
Where is Abuse Concentrated in New TLDs?

Exploited New TLDs **MAY 2017**	Abuse Domains Reported to DAAR	New TLD Program Resolving Domains for Which DAAR Obtains Data
5 most exploited new TLDs	56%	22%
10 most exploited new TLDs	73%	34%
25 most exploited new TLDs	97%	70%

Exploited New TLDs **SEP 2017**	Abuse Domains Reported to DAAR	New TLD Program Resolving Domains for Which DAAR Obtains Data
5 most exploited new TLDs	53%	26%
10 most exploited new TLDs	71%	48%
25 most exploited new TLDs	95%	67%

TLDs for which no abuse domains were reported are not included in the counts

Project Status



Project Status

- Doing it right is more important than doing it fast
  - Reviewing our data feeds and licensing
  - Tuning collection systems to ensure timely and resilient updates
  - Third party independent review of our methodology
- Version 2.0 features under development
  - Additional automation for reporting
  - Granular attribution
  - Experimentation with additional measurements

https://www.flickr.com/photos/artisticbokeh/

Exploring the feasibility of…

- Delving deeper into how domain names are used
  - Legitimate vs. compromised domain data?
  - Tracking recidivism?
  - Additional classifications of spam domains
  - Add URL amplification: how many unique abuse URLs are associated with a unique abuse domain?
- Including more data?
  - IP and ASN reputation data?
  - Registration fee data
  - New (additional) reputation data

https://www.flickr.com/photos/artisticbokeh/

Questions?

Thank You

@icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

slideshare/icannpresentations

soundcloud/icann

Dave's Contact Info:

dave.piscitello@icann.org
@securityskeptic
www.securityskeptic.com
about.me/davepiscitello