ICANN 63

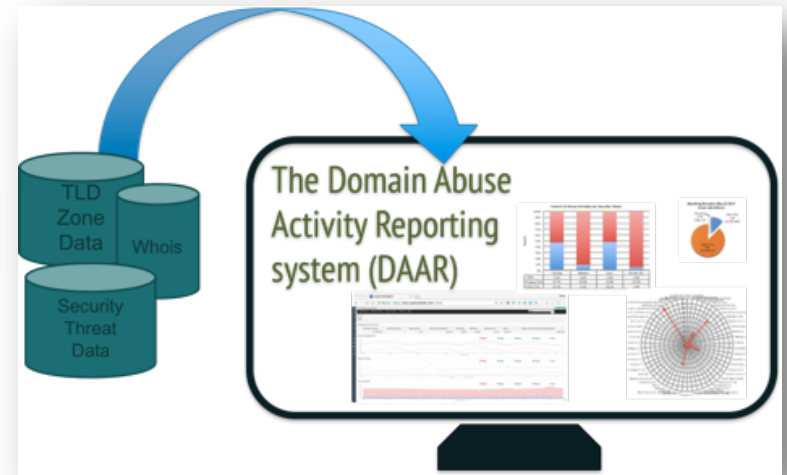## ANNUAL GENERAL

BARCELONA

20–25 October 2018

The Domain Abuse Activity Reporting System (DAAR)

John Crain

Barcelona
October 2018

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 Program

- ◉ Goal of Open Data Program 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 Program

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

```
                                    🏠 dave.piscitello —
Domain Name: GOOGLE.COM
Registry Domain ID: 2138514_DOMAIN_COM-VRSN
Registrar WHOIS Server: whois.markmonitor.com
Registrar URL: http://www.markmonitor.com
Updated Date: 2011-07-20T16:55:31Z
Creation Date: 1997-09-15T04:00:00Z
Registry Expiry Date: 2020-09-14T04:00:00Z
Registrar: MarkMonitor Inc.
Registrar IANA ID: 292
Registrar Abuse Contact Email: abusecomplaints@m
Registrar Abuse Contact Phone: +1.2083895740
Domain Status: clientDeleteProhibited https://ic
Domain Status: clientTransferProhibited https://
Domain Status: clientUpdateProhibited https://ic
Domain Status: serverDeleteProhibited https://ic
Domain Status: serverTransferProhibited https://
Domain Status: serverUpdateProhibited https://ic
Name Server: NS1.GOOGLE.COM
Name Server: NS2.GOOGLE.COM
Name Server: NS3.GOOGLE.COM
Name Server: NS4.GOOGLE.COM
```

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
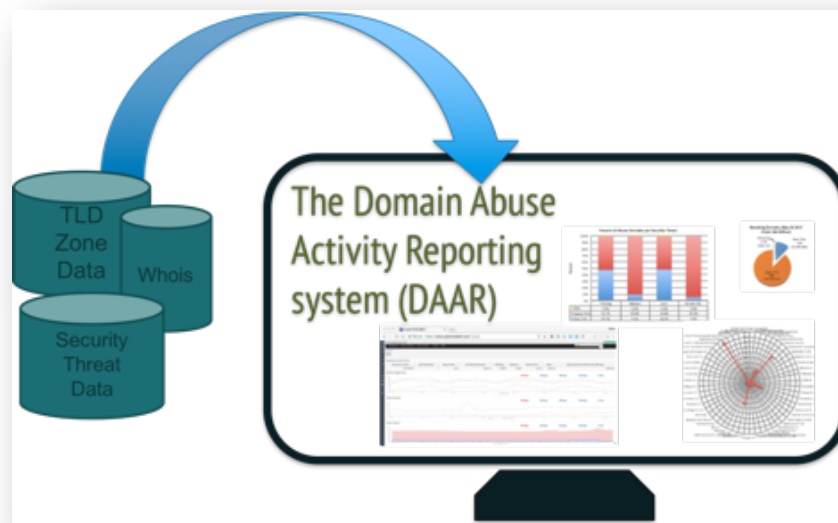
DAAR Is Not An Abuse List Service

- ⊙ ICANN 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 Threat Exchange 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

Does DAAR Identify All Abuse?

No:

DAAR lists domain names associated with abuse identified by third parties.

Only those names associated with generic TLDs are measured and only for specific abuse types.

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

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

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

Project Status

The Domain Abuse Activity Reporting system (DAAR)

TLD Zone Data

Whois

Security Threat Data

New Hire!!
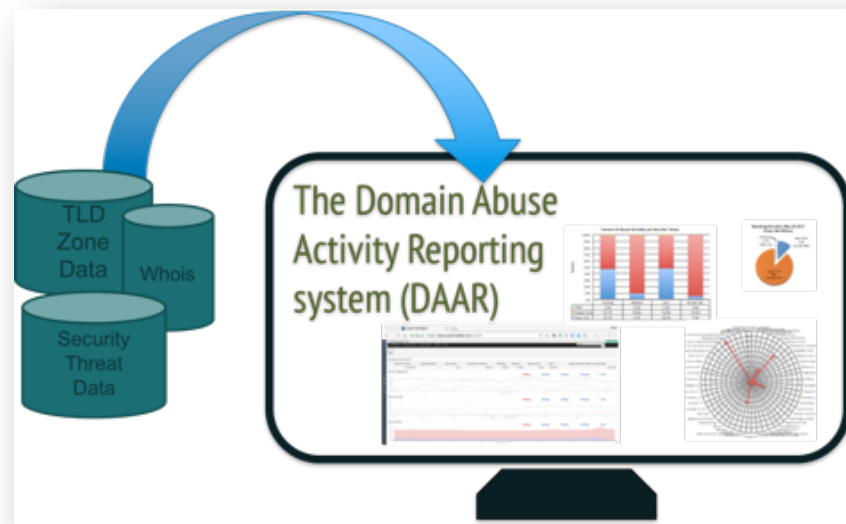
Samaneh Tajalizadehkhoob joined ICANN as of Oct 1

Samaneh holds a PHD degree in Information Security and Data Analytics from the Delft University of Technology in the Netherlands. She worked as a Post-Doctoral researcher at the same University and was visiting scholar at KU Leuven, DistriNet Research Group, where she was carrying out internet measurements to estimate web vulnerabilities. Samaneh has publications on WebSecurity, Cyber Security and Malware.

Her work at ICANN will focus on SSR research, including the DAAR project.

\

Project Status

○ Methodology Review of the Domain Abuse Activity Reporting (DAAR) System

- Two experts asked to review methodology
- Methodology and reviews posted for comment

○ Comments to DAAR@icann.org

○ https://www.icann.org/news/announcement-2018-07-20-en

Project Status and Next Steps

o Five comments received

o Ironically hundreds of spam messages, and a few phish.

We are currently assessing the received comments and intend to publish answers to each comment or question in written form by December 1, 2018

\

Project Status and Next Steps

o Publishing Monthly Reports

- As of 2019 ICANN will be publishing monthly reports from the DAAR system.

These will contain data similar to that which I will show tomorrow for the month including deltas.

We are still investigating publication of data into the Open Data Program

\

Project Status and Next Steps

o Discussions with Registries who are interested in viewing their own data

The SSR team does this in the context of sharing and learning from other security professionals in the Industry.

The data has already enabled constructive and data driven discussions with Industry members.

\

Visualizing DAAR Data

The Domain Abuse Activity Reporting system (DAAR)

TLD Zone Data

Whois

Security Threat Data
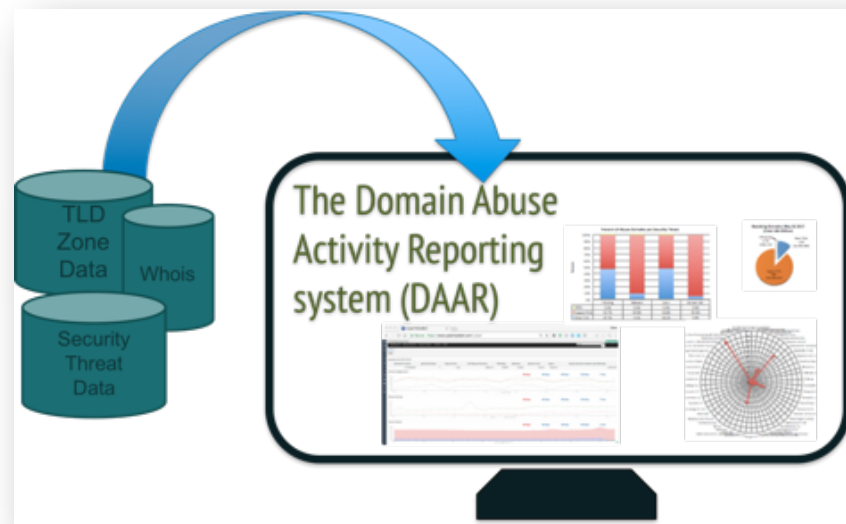
Sept 30 2018 Snapshot: Phishing Domains Percent of Abuse

Phish Percent
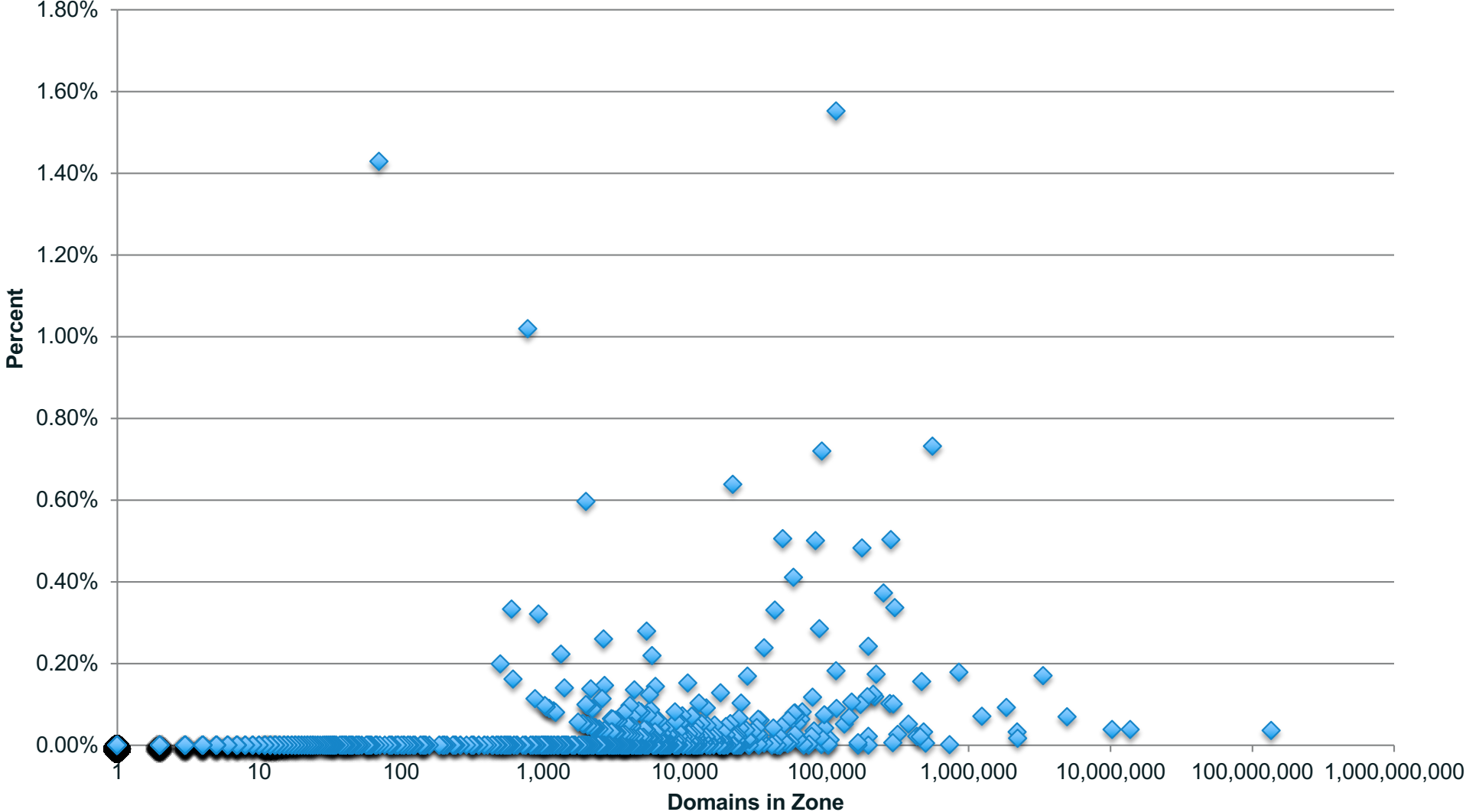
Sept 30 2018 Snapshot: Malware Domains Percent of Abuse

Malware Percentage
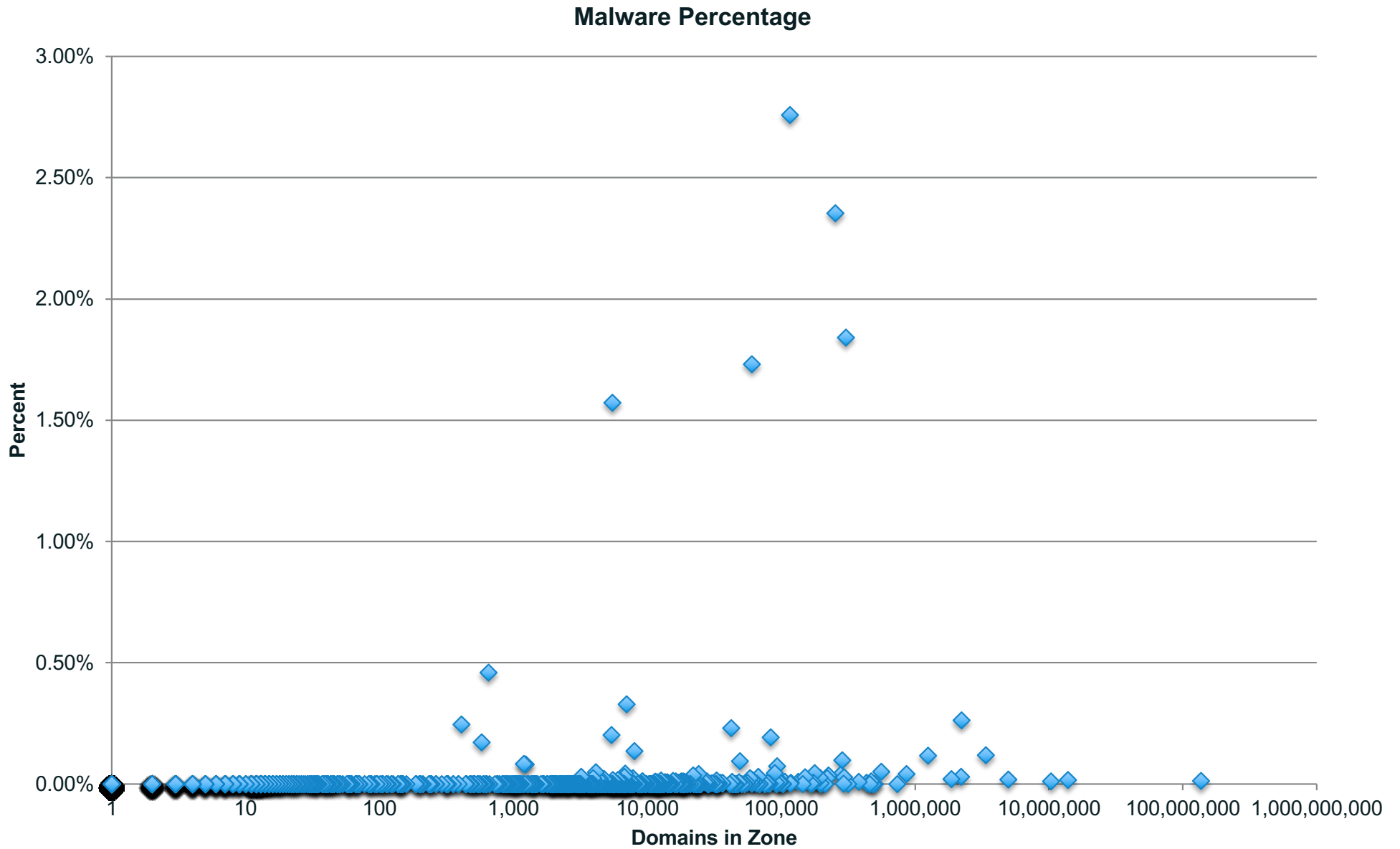
Sept 30 2018 Snapshots: Spam Domains Percent of Abuse

Spam Percent
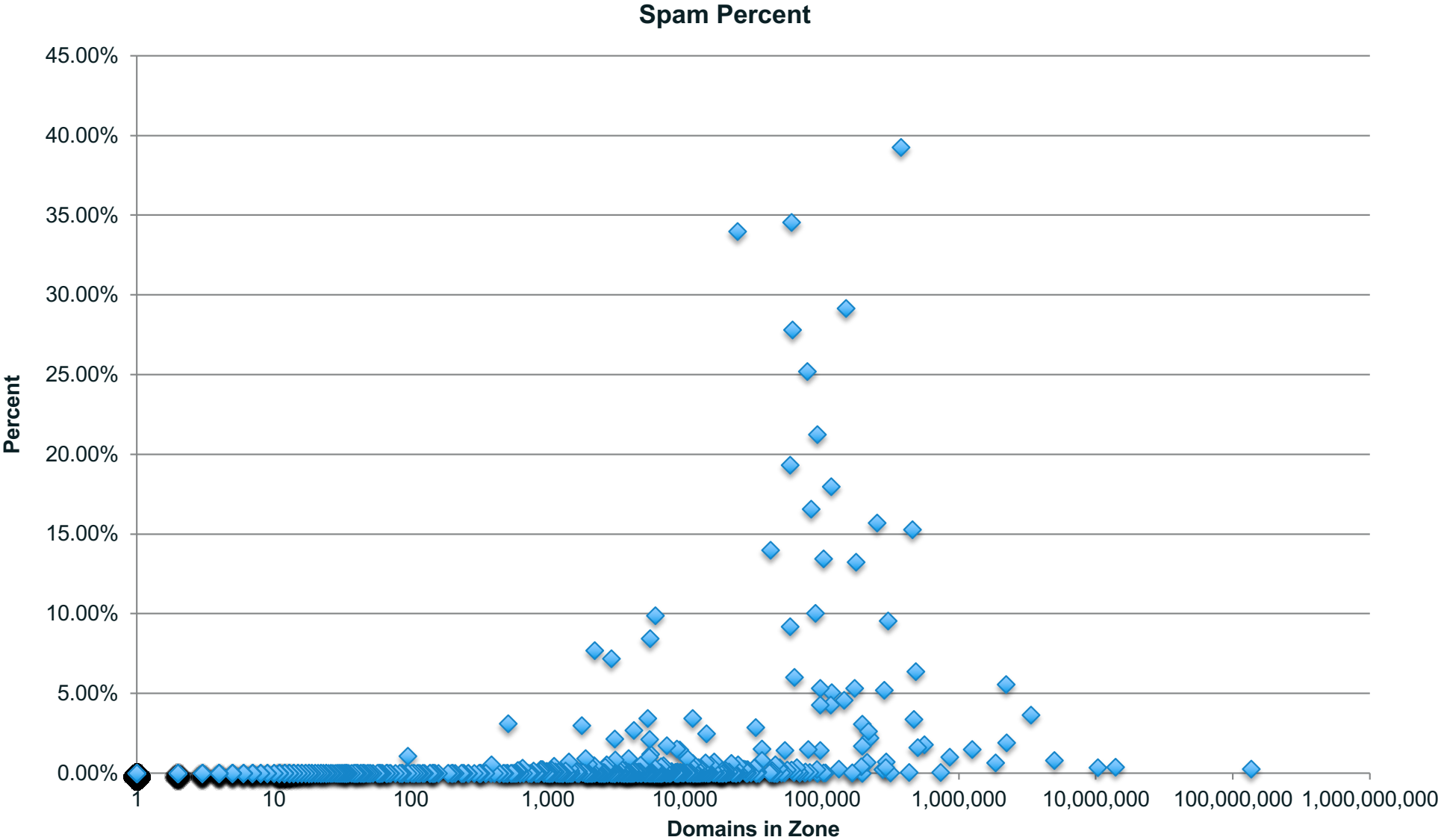
Sept 30 2018 Snapshots: Abuse Domains Percent of Abuse
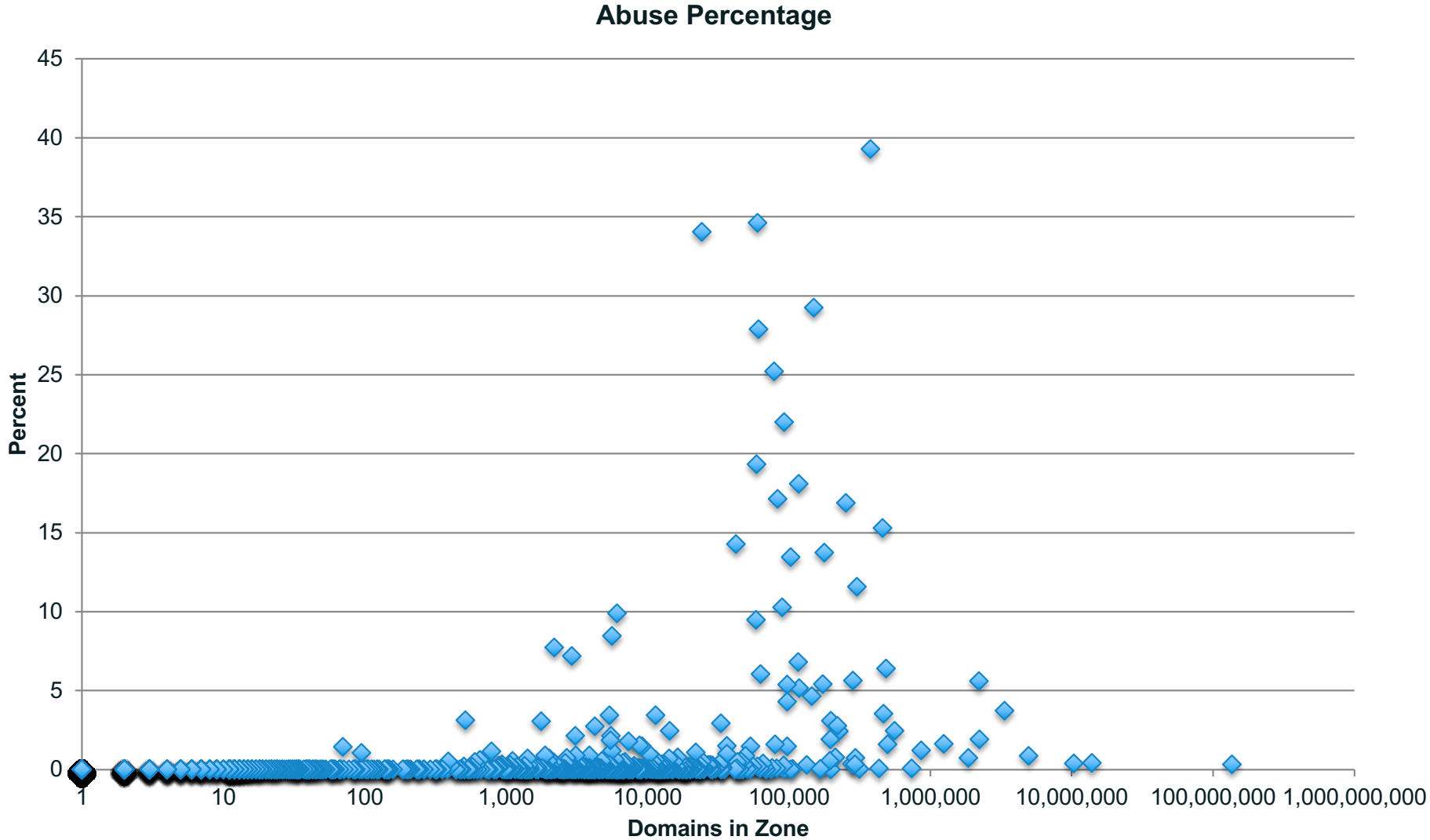
Abuse Percentage

Ranked by percentage

Number of TLDS	Percentage of Abuse Domains
Top 5	15%
Top 10	20%
Top 25	36%

Once you get beyond the top 65 the TLDs have an abuse percentage of 1% in their portfolios

Domains Resolving
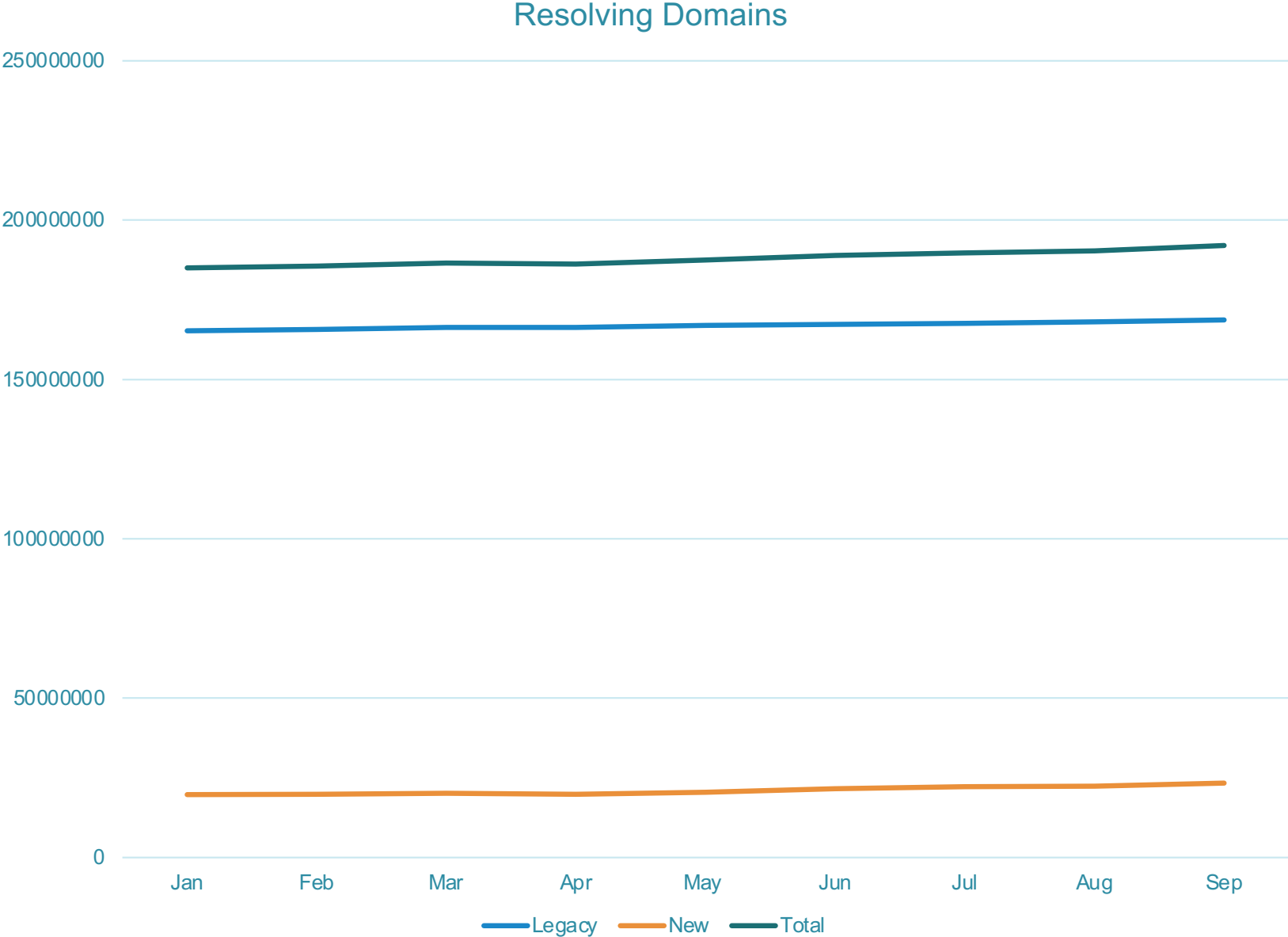
Resolving Domains

The number of resolving domains is still increasing

gTLD have seen an increase of 7 Million resolving domains since January

+/- 168.5 Million Domains in the older TLDs

+/- 23.5   Million Domains in the New TLDs

ICANN

Unique Abuse Domains

Unique Abuse Domains

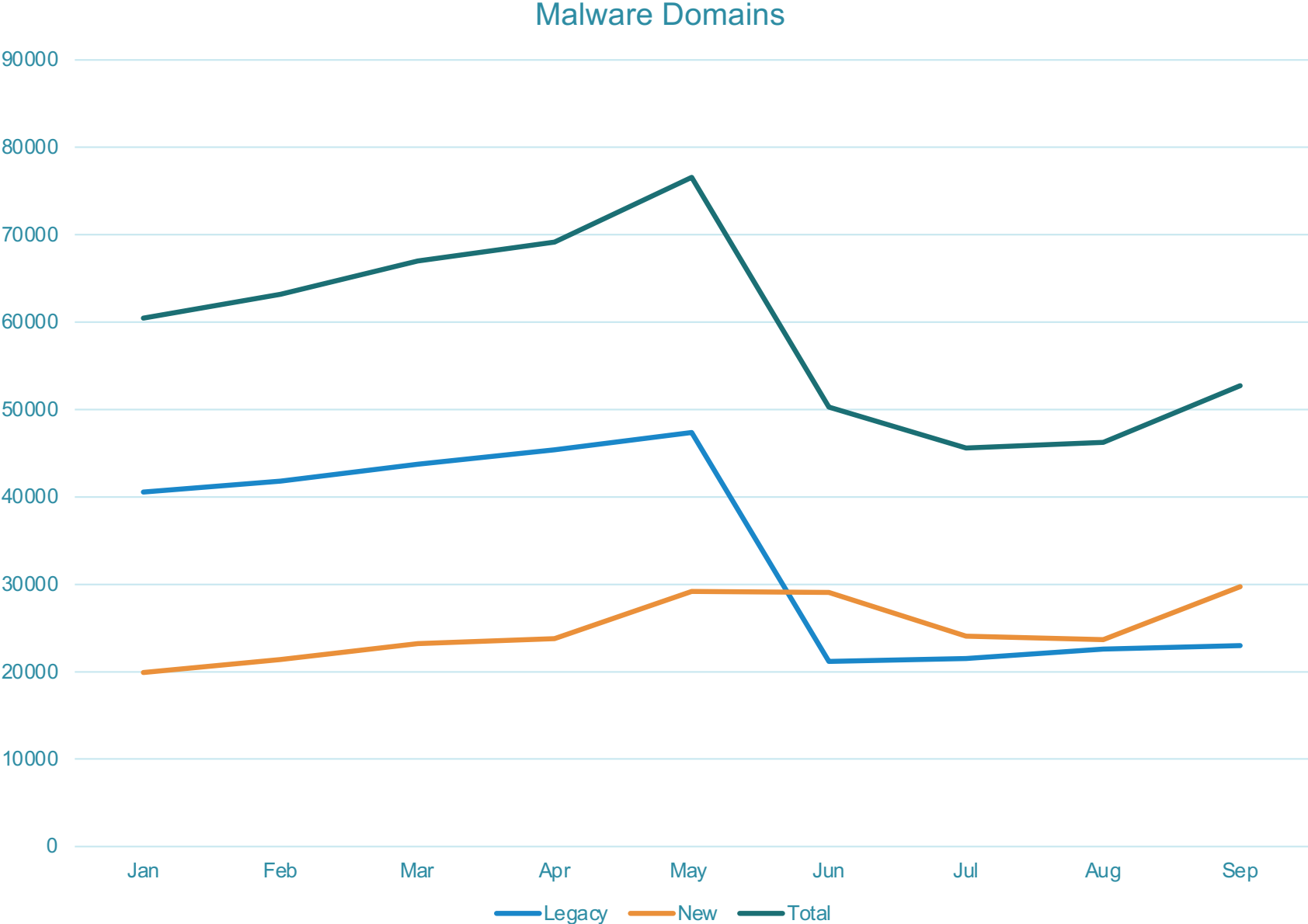gTLD have seen a decrease of about 0.5 Million listed abuse domains domains since a February spike

+/- 700,000 Domains in the older TLDs
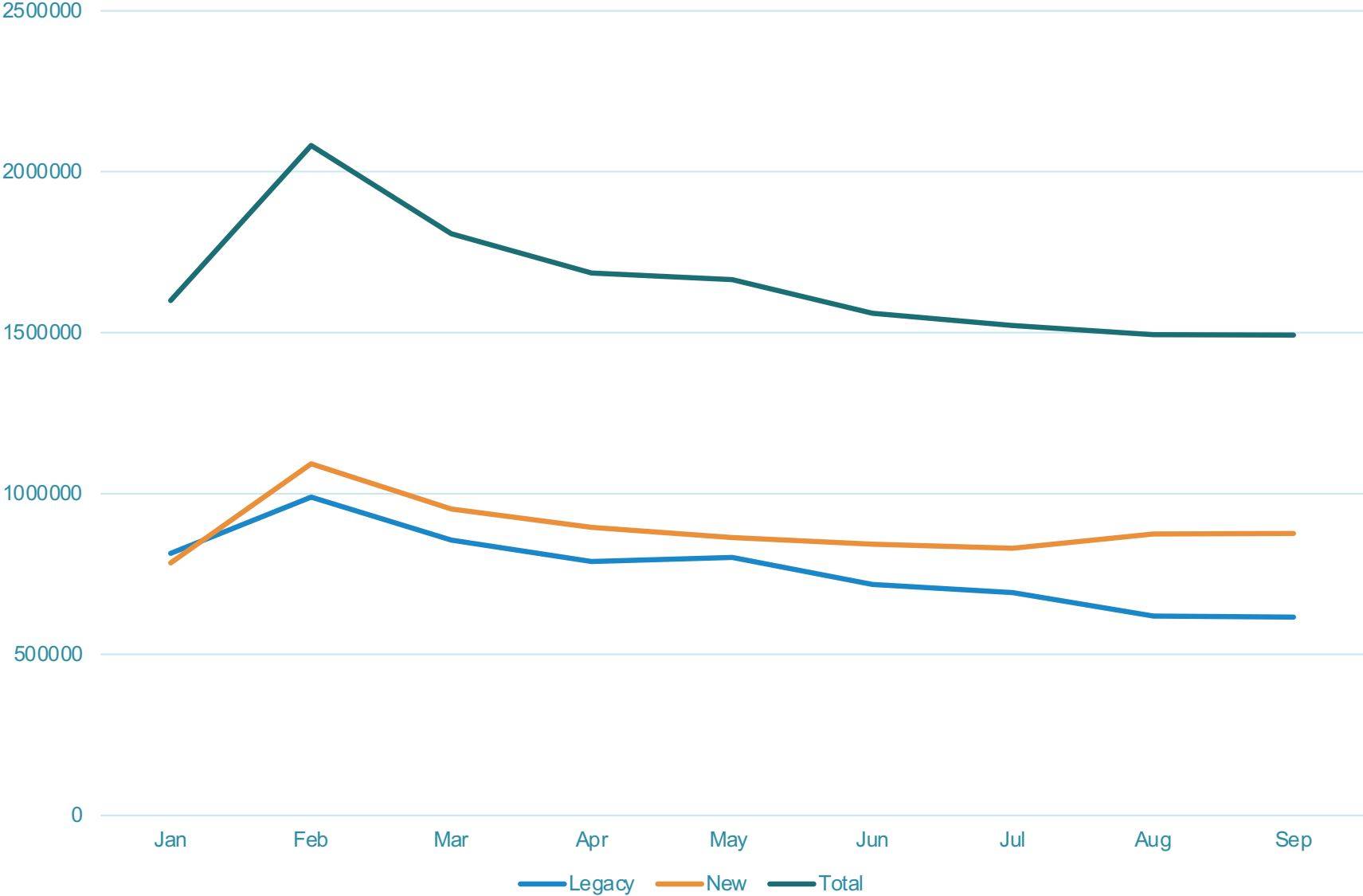
+/- 900,000 Domains in the New TLDs
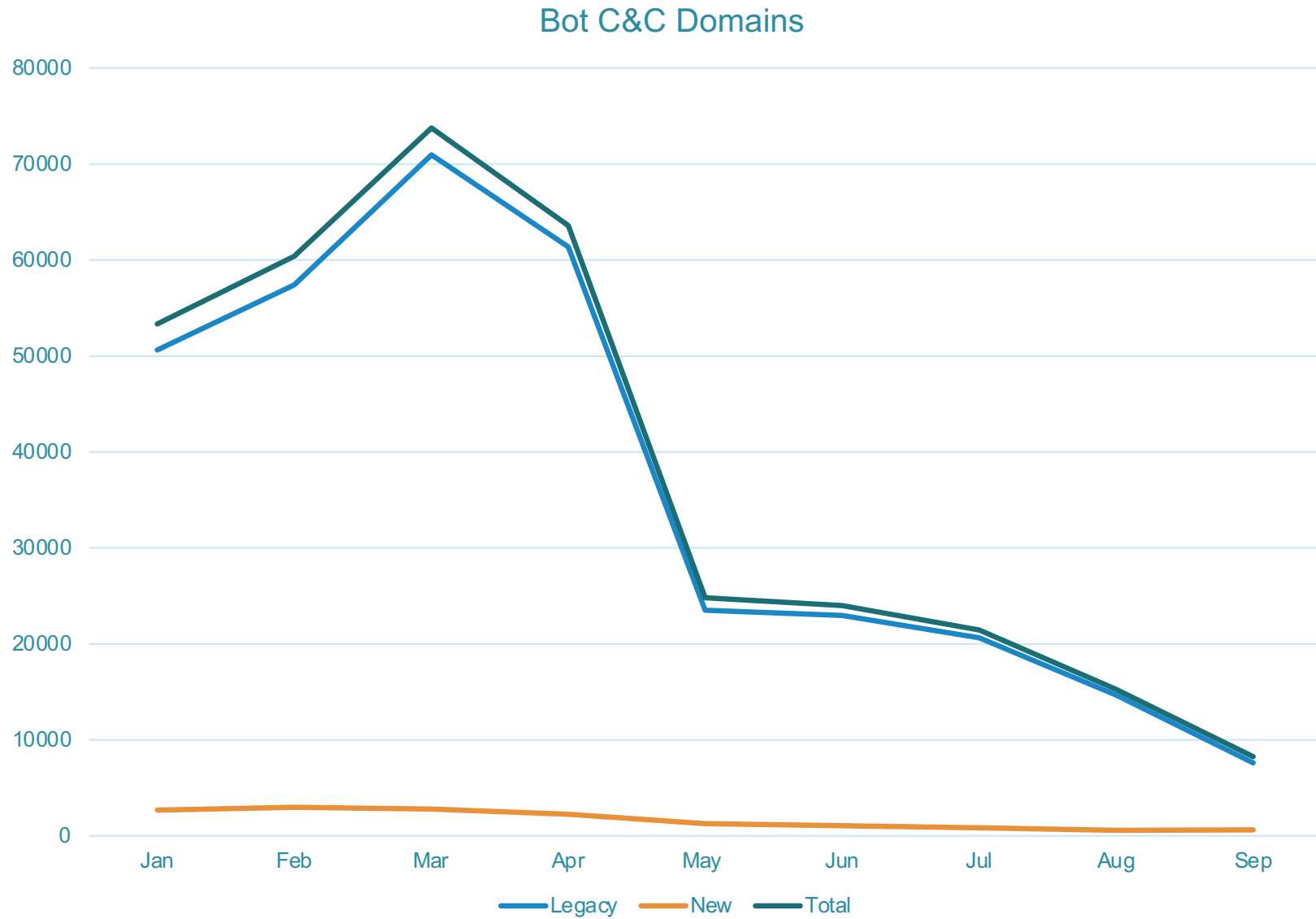
Phish Domains

Phish Domains

Legacy — New — Total
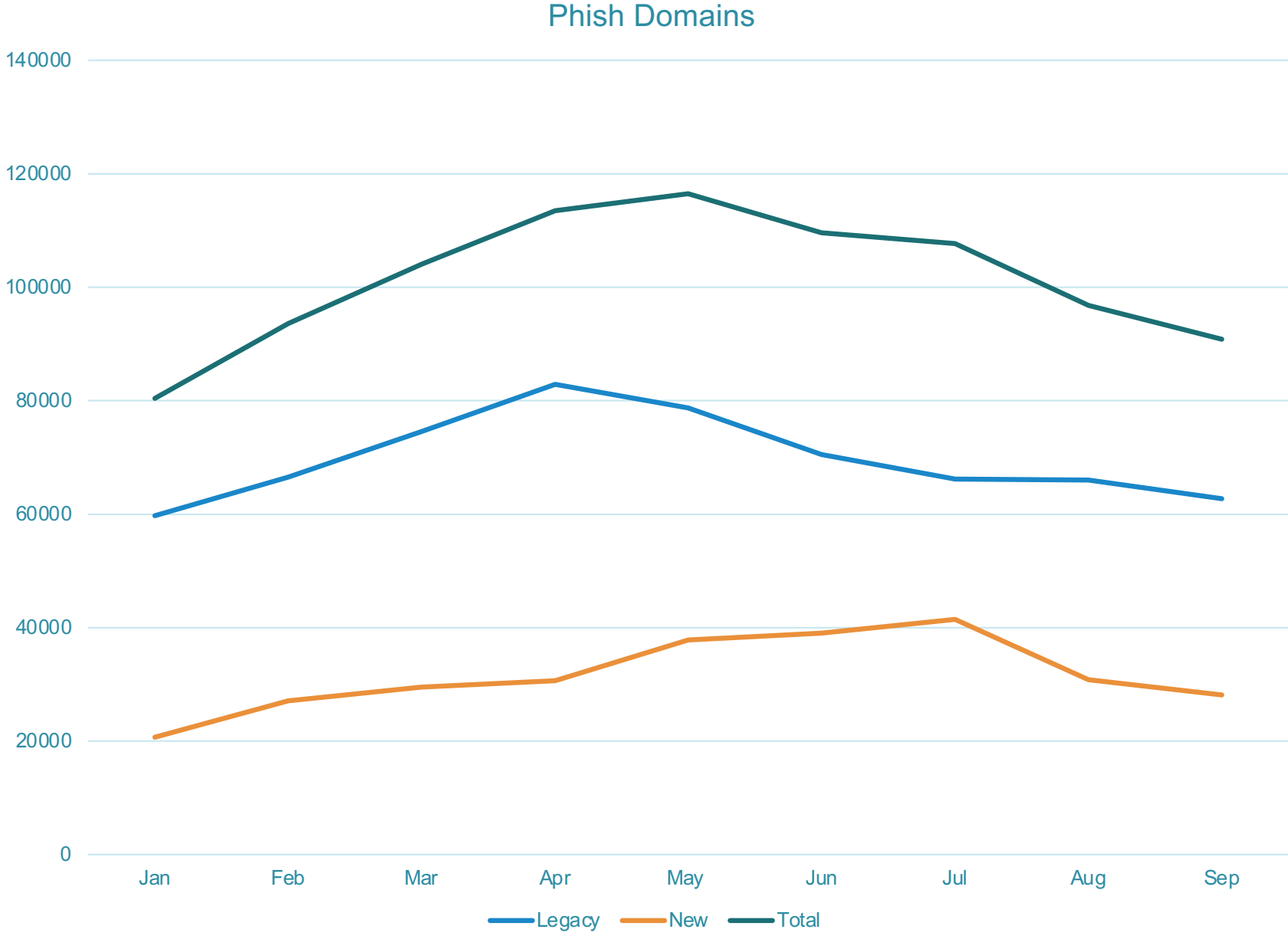
Malware Domains

Malware Domains

Spam Domains

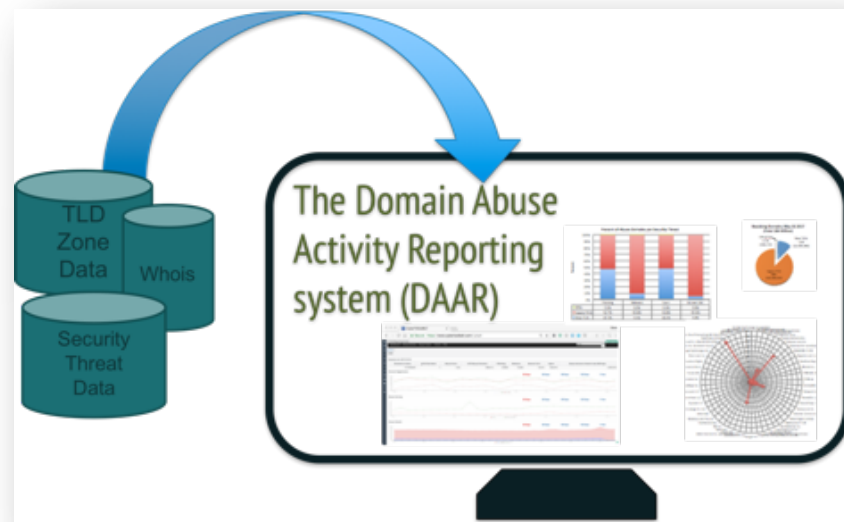Spam Domains

BotNet C&C Domains
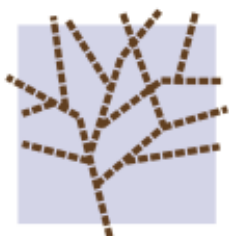
Bot C&C Domains

Legend: Legacy — New — Total

Where do we want to go from here?

Measuring abuse

o We are always open to discussion on improvements or other ways the data can be used to help inform discussions around DNS abuse

o Feel free to use daar@icann.org to contact us

\

Discussions on DNS Abuse at IDS (May 10-11)

Questions?

Thank You

[Twitter] @icann

[Facebook] facebook.com/icannorg

[YouTube] youtube.com/icannnews

[Flickr] flickr.com/icann

[LinkedIn] linkedin/company/icann

[LinkedIn] slideshare/icannpresentations

[SoundCloud] soundcloud/icann

John's Contact Info:
   John.crain@icann.org
Twitter: @johnlcrain
Skype: JohnLcrain

DAAR@icann.org