# DNS Research and Analytics

**SIDN Connect**

Roy Arends

November 2019

# What are root-servers ?

- Root servers are authoritative servers responsible for answering DNS questions about the root zone.
- 12 Organizations are responsible for 13 hosts
  - ICANN is one of them.
- 13 hosts are deployed over hundreds of locations
- Using Anycast, the number of instances > 1000
- The IANA function is responsible for root-zone management.
  - IANA function is run by PTI, an affiliate of ICANN

# What do we see at our root-server

- We see more and more traffic at our root-server
- The root-server system acts as a filter for the rest of the domain name system.
- when a question is syntactically correct and the answer is known, only then a response is “useful”.
- When SIDN (.NL) gets a question from a resolver, it is highly likely that the root-server system was asked before.
- So what do we see at root-servers?

# Wow, That's a Lot of Packets

Duane Wessels, Marina Fomenkov

*Abstract*— **Organizations operating Root DNS servers report loads exceeding 100 million queries per day. Given the design goals of the DNS, and what we know about today's Internet, this number is about two orders of magnitude more than we would expect.**

**With the assistance of one root server operator, we took a 24-hour trace of queries arriving at one of the thirteen root servers. In this paper we analyze these data and use a simple model of the DNS to classify each query into one of nine categories. We find that, by far, most of the queries are repeats and that only a small percentage are legitimate.**

**We also characterize a few of the "root server abusers," that is, clients sending a particularly large number of queries to the root server. We believe that much of the root server abuse occurs because the querying agents never receive the replies, due either to packet filters, or to routing issues.**

*Keywords*— **DNS root server**

## I. Background: *DNS 101*

The Domain Name System (DNS) is a fundamental component of the modern Internet [1], providing a critical link between human users and Internet routing infrastructure by mapping host names to IP addresses. The DNS utilizes a hierarchical name space divided into zones, or domains. This hierarchy is manifested in the widespread "dots" structure. For example, `com` is the parent zone for `example.com`, `microsoft.com`, `cnn.com`, and approximately 20 million other zones.

Each zone has one or more authoritative name servers. These are dedicated servers, whose job is to answer queries for names within their zone(s). For example, UCSD has three authoritative name servers. An application that needs to know the IP address for `www.ucsd.edu` can send a DNS query to one of those servers, which then returns an authoritative answer. If the application does not know where to send a query it asks the servers in the parent zone. In the example above, not knowing anything about `ucsd.edu`, the application should send a query to the authoritative server for the `edu` zone. If the application does not know about the `edu` zone, it queries the "root zone." This process is called *recursive iteration*.

The DNS root zone is served by 13 name servers (not to be confused with the 13 generic top-level domain servers) distributed across the globe. Thirteen is the maximum number of root servers possible in the current DNS architecture because that is the most that can fit inside a 512-byte UDP reply packet. Ten root servers are located in the U.S., two are in Europe, and one is in Asia.[1] The root zone and the root name servers are vital because they are the starting points for locating anything in the DNS. Without them, the DNS and hence almost every application we use (the Web, ssh, email) would be rendered unusable.

DNS clients, or resolvers, that query name servers, come in one of two flavors: stub and recursive. Stub resolvers, typically found in user applications, such as web browsers, ssh clients, and mail transfer agents, are rather primitive and mostly rely on smarter recursive resolvers that understand name server referrals. Recursive resolvers are usually implemented in specialized DNS applications such as the Berkeley Internet Domain Name (BIND) [2] server and Microsoft's DNS server. Most organizations operate local recursive name servers.

Recursive name servers cache name server responses, including referrals. Caching conserves network resources because intermediate servers do not need to query the root name servers for every request. For example, the name server learns that `a.gtld-servers.net` and others are authoritative for the `com` zone and sets the time-to-live (TTL) for this information. Typical TTLs for top level domains are on the order of 1–2 days.

In theory, a caching recursive name server only needs to query the root name servers for an unknown top level domain or when a TTL expires. However, a number of studies have shown that the root name servers receive many more queries than they should. In this paper we thoroughly investigate and characterize root name server traf-

The Measurement Factory, Inc., Boulder, Colorado, E-mail: `wessels@measurement-factory.com`.

CAIDA, San Diego Supercomputer Center, University of California, San Diego. E-mail: `marina@caida.org`.

Support for this work is provided by WIDE and DARPA NMS N66001-01-1-8909.

[1] In fact many of the root name servers are actually multiple hosts behind network load balancers. Some of them even occupy a few physical locations, employing IPv4 anycast to operate under a single IP address.

# Wow, That's a Lot of Packets

Duane Wessels, Marina Fomenkov

| Type | Count | Percent |
|---|---|---|
| Unused Query Class | 36,313 | .024 |
| A for A | 10,739,857 | 7.03 |
| Unknown TLD | 19,165,840 | 12.5 |
| Nonprintable in query | 2,962,471 | 1.94 |
| RFC1918 PTR | 2,452,806 | 1.61 |
| Identical Query | 38,838,688 | 25.4 |
| Repeated Query | 68,610,091 | 44.9 |
| Referral Not Cached | 6,653,690 | 4.36 |
| **Legitimate** | **3,284,569** | **2.15** |

TABLE II

QUERY CLASSIFICATION RESULTS (24-HOUR PERIOD ON 4 OCTOBER 2002 AT THE F-ROOT DNS SERVER).

more queries than they should. In this paper we thoroughly investigate and characterize root name server traf-

[1]In fact many of the root name servers are actually multiple hosts behind network load balancers. Some of them even occupy a few physical locations, employing IPv4 anycast to operate under a single IP address.

# What do we see at our root-server

- IMRS statistics for 18th Nov 2019
  - total: 12.9 G responses (12.859.473.447)
    - 81.44% UDP-v4
    - 14.31% UDP-v6
    - 3.59% TCP-v4
    - 0.66 % TCP-v6
- 98.5% queries saw a response.
  - High-frequency identical queries get one response
- We’re going to ignore TCP for this effort (not statistically significant)

# What do we see at our root-server

| AA | RCODE | Ans | Description | |
|---|---|---|---|---|
| Set | NOERROR | 0 | NODATA | 0.84% |
| Set | NOERROR | 1+ | Auth Ans. | 3.31% |
| Set | NXDOMAIN | 0 | NXDOMAIN | 62% |
| Clear | NOERROR | 0 | Delegation | 34% |

# What about caching

- 34% of all queries result in delegations
- All delegation point NS records have a 2-day TTL
- Proper caching: at most 1 query per TLD per source IP
- Of the 34% delegation responses:
  - 98 % are duplicates
  - 2% are unique

# What does bogus look like

- 2.7% of Authoritative NODATA are for type A6
- Large amount of proper delegations are for RFC1918 reverse address space (and other lame addresses)
- Reflection and Amplification attacks
- Spam traffic (loads of MX queries)
- DGA related traffic

# Conclusion

- The root server system is One Big Filter for loads of bad queries
  - Only 34% result in a delegation
- The bulk of the 62% should never have been send in the first place
- The bulk of the 34% should have been properly cached.
- The 34% of delegations still contains loads of DGA, RFC1918 address space, spam traffic.
- It is nearly impossible to “fix” any of this “at the root”
  - (if you don’t respond, things get worse)
- Some recommendations for resolvers:
  - Properly cache, local root copy, ACLs, domain block lists

# ITHI: an ICANN Initiative

- **ITHI**, or **I**dentifier **T**echnologies **H**ealth **I**ndicators is an ICANN initiative to "**measure**" the "**health**" of the "**identifier system**" that "**ICANN helps coordinate**".

- The goal is to produce a set of **indicators** that will be **measured and tracked over time** that will help determine if the system of identifiers is overall doing better or worse.

- This is a long-term project, expected to run for a number of years.

# ITHI Phases

- **Phase 1: Analysis (2015-2016)**
  - Strategic choice to define problem areas first
  - Many discussions with the larger community
  - Split of project ICANN / RIR
- **Phase 2: Development (2017-2018)**
  - Building platform
  - Finding partners
  - Getting data
- **Phase 3: Sustaining (2019-…)**

We are here now

# Simplified Indicator Dashboard

Home Metrics Participate About

| ITHI by ICANN | | Full table |
|---|---|---|
| **Identifier Technology Health Indicator** | | **As of Nov 2019** |
| % No Such Domain queries seen by root servers | | 74.60% |
| % of resolvers that perform DNSSEC validation | | 33.33% |
| %requests to top name at the root | .LOCAL | 3.56% |
| %requests to top name at resolvers | .UNIFI | 0.04% |
| Number of resolvers seeing 50% of first queries | | 212 |
| Phishing Domains per 10,000 registered names | | 2.08 |

*The home page at ithi.research.icann.org provides a quick view of chosen indicators.*

# Complete Dashboard

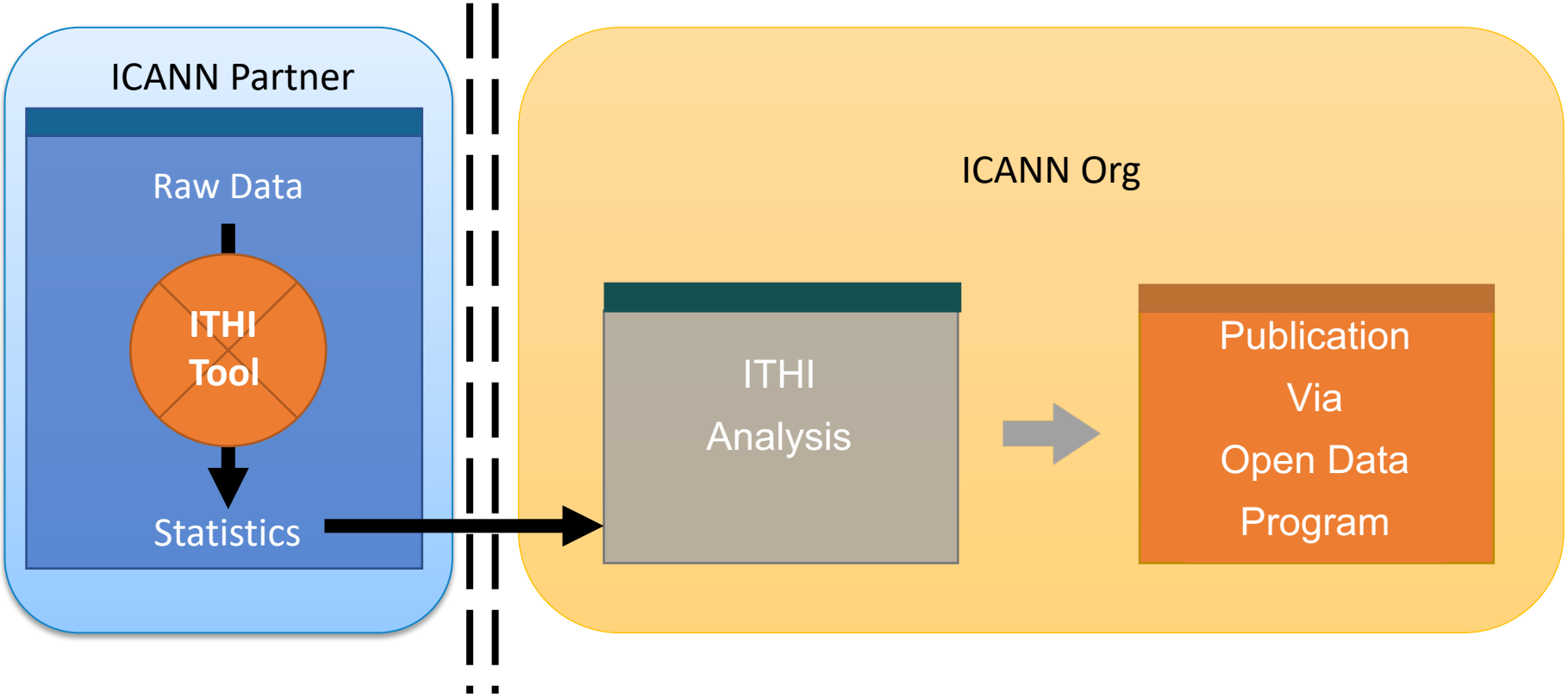
| ITHI by ICANN | Identifier Technology Health Indicator | | As of Nov 2019 | Past 3 months | Historic Low | Historic High |
|---|---|---|---|---|---|---|
| Root Server Health | % No Such Domain queries seen by root servers | | 74.60% | 74.93% | 62.95% | 75.10% |
| DNSSEC Deployment | % of resolvers that perform DNSSEC validation | | 33.33% | 32.28% | 23.43% | 32.33% |
| Name collision | %requests to top 3 names at the root | .LOCAL | 3.56% | 3.33% | 2.36% | 4.47% |
| | | .HOME | 2.82% | 2.58% | 2.48% | 3.67% |
| | | .LAN | 1.20% | 0.98% | 0.47% | 1.05% |
| | %requests to top 3 names at resolvers | .UNIFI | 0.04% | 0.06% | 0.03% | 0.09% |
| | | .DNS | 0.02% | 0.02% | 0.00% | 0.03% |
| | | .LOCAL | 0.01% | 0.02% | 0.00% | 0.06% |
| Resolver Concentration | Number of resolvers seeing 50% of first queries | | 212 | 208.85 | 205.50 | 212.19 |
| | Number of resolvers seeing 90% of first queries | | 2180 | 2094.85 | 2036.90 | 2152.81 |
| Dns Abuse (as of Aug 2019, measured on 1193 GTLD and 1793 registrars) | Abuse Domains per 10,000 registered names | Phishing | 2.08 | 2.72 | 2.43 | 4.13 |
| | | Malware | 1.16 | 1.11 | 1.10 | 2.00 |
| | | Botnets C&C | 0.53 | 0.37 | 0.54 | 1.48 |
| | | Spam | 16.27 | 14.70 | 56.56 | 61.89 |
| | Number of GTLD to account for 50% of abuses | Phishing | 1 | 1.67 | 1 | 3 |
| | | Malware | 1 | 1.00 | 1 | 3 |
| | | Botnets C&C | 3 | 2.33 | 2 | 3 |
| | | Spam | 3 | 3.33 | 4 | 5 |
| | Number of GTLD to account for 90% of abuses | Phishing | 9 | 12.33 | 11 | 19 |
| | | Malware | 7 | 7.33 | 7 | 19 |
| | | Botnets C&C | 5 | 5.00 | 4 | 5 |
| | | Spam | 22 | 24.33 | 18 | 28 |

| Metric | Name | Data Source |
|---|---|---|
| M1: | Inaccuracy of Whois Data | ICANN compliance dept. |
| M2: | Domain Name Abuse | ICANN's DAAR Project https://www.icann.org/octo-ssr/daar |
| M3: | DNS Root Traffic Analysis | Samples of DNS root traffic |
| M4: | DNS Recursive Server Analysis | Summaries of recursive resolvers traffic |
| M5: | DNS Resolver Behavior | APNIC |
| M6: | IANA registries for DNS parameters | Scan of recursive resolvers traffic |
| M7: | DNSSEC Deployment | Snapshots of DNS root zone |
| M8: | DNS TLD Traffic Analysis | Summaries of TLD traffic |

# ITHI Data: ICANN + Partners + Contracts

- **ICANN (Internal Data)**
  - Compliance department (M1)
  - DAAR (M2)
  - L-Root data (M3)
  - Root zone (M7)
- **White box measurements with partners**
  - Measurements at recursive & authoritative servers
  - M4, M6, M8
- **Black box measurements**
  - APNIC/Google Ads platform
  - Eyeball view of resolvers M5

# Privacy




No PII, only statistics, are sent to ICANN org

No “naming and shaming”

# DAAR: Domain Abuse Activity Reporting

"Systems are particularly prone to failure when the person guarding them is not the person who suffers when they fail." (Ross Anderson, 2001)

Lack of security is an incentive problem as much as it is a technical problem.

A growing need for proactive detection and mitigation strategies by actors that operate domain names.

There is lack of knowledge about security threat concentrations in TLDs and their operators.

# What is DAAR?

- A system for reporting on domain name registration and security threat data across TLD registries
- DAAR data can be used to
  - Report on threat activity at TLD level
  - Study historical security threats or domain registration activity
  - Help operators understand or consider how to manage their reputations, anti-abuse programs, or terms of service
  - More informed security decision making and policy

# Where does the data come from?

- DNS zone data
  - Publicly available methods Centralized Zone Data Service (CZDS) 1220 gTLDs, 192 million domains
- Published WHOIS registration data
  - Accurate registrar reporting depends on WHOIS Scaling data collection
- Open source data
- commercial abuse threat data
- reputation blacklist (RBL) data
- Some of these data feeds require a license or subscription

# Where does the data come from?

- DAAR uses multiple abuse Reputation Blocklist (RBL) datasets to generate
  - Daily raw counts of domains associated with security threat
  - Daily total and cumulative percentage security threat domains
  - Calculate monthly/yearly newly added security threat domains
  - Visual analytics regarding security threat trends
- DAAR collects domain data for
  - Phishing
  - Malware
  - Spam
  - Botnet Command & Control

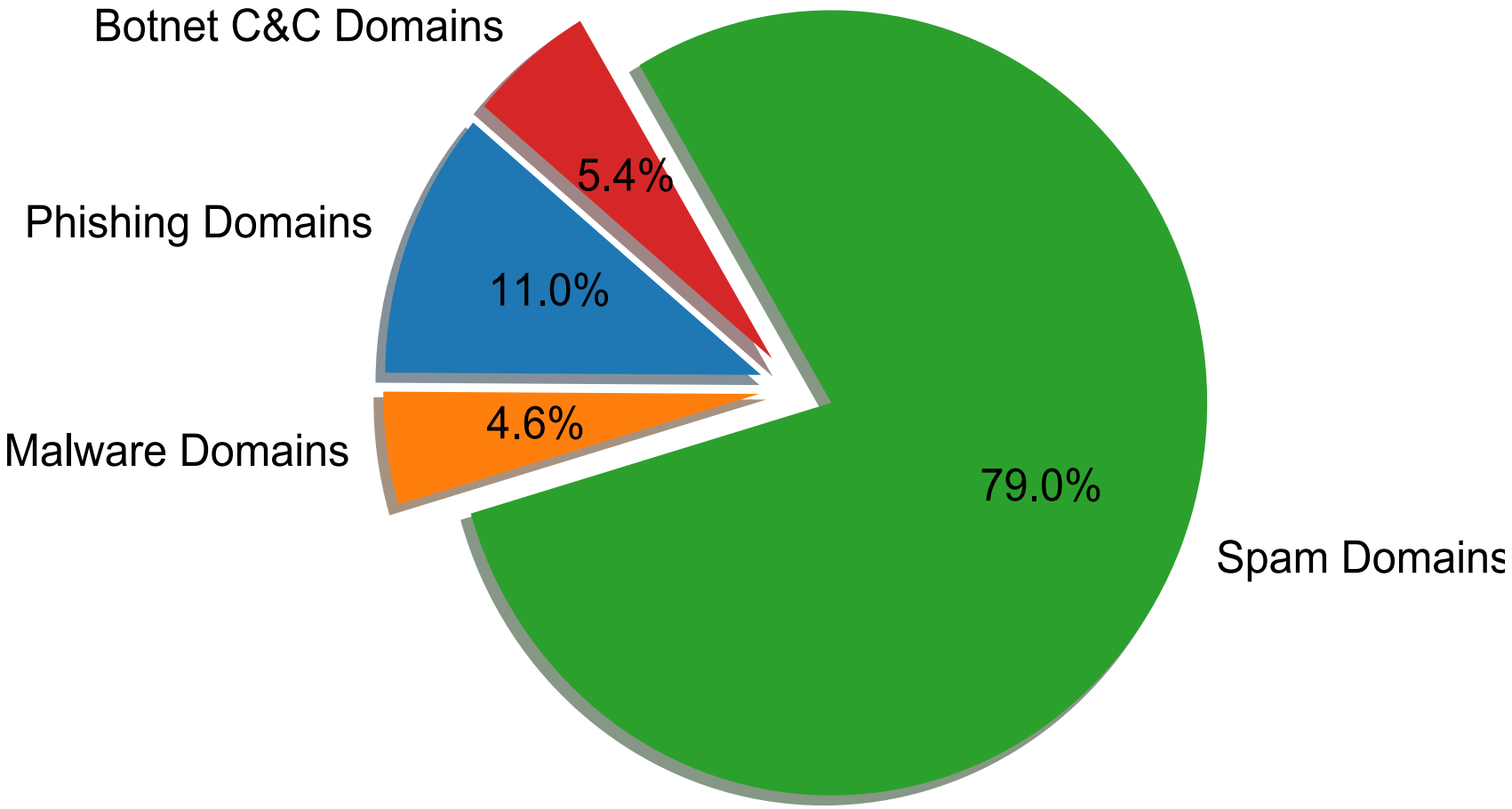
# Where does the data come from?

- SURBL lists
  - (Spam – Phishing - Malware)
- Spamhaus Domain Block List
  - (Spam - Phishing - Malware - Botnet C&C)
- Anti-Phishing Working Group
  - (Phishing)
- Malware Patrol
  - (Malware, Ransomware, Botnet C&C )
- Phishtank
  - (Phishing domains)
- ABUSE.CH
  - (Ransomware tracker, Feodo tracker)

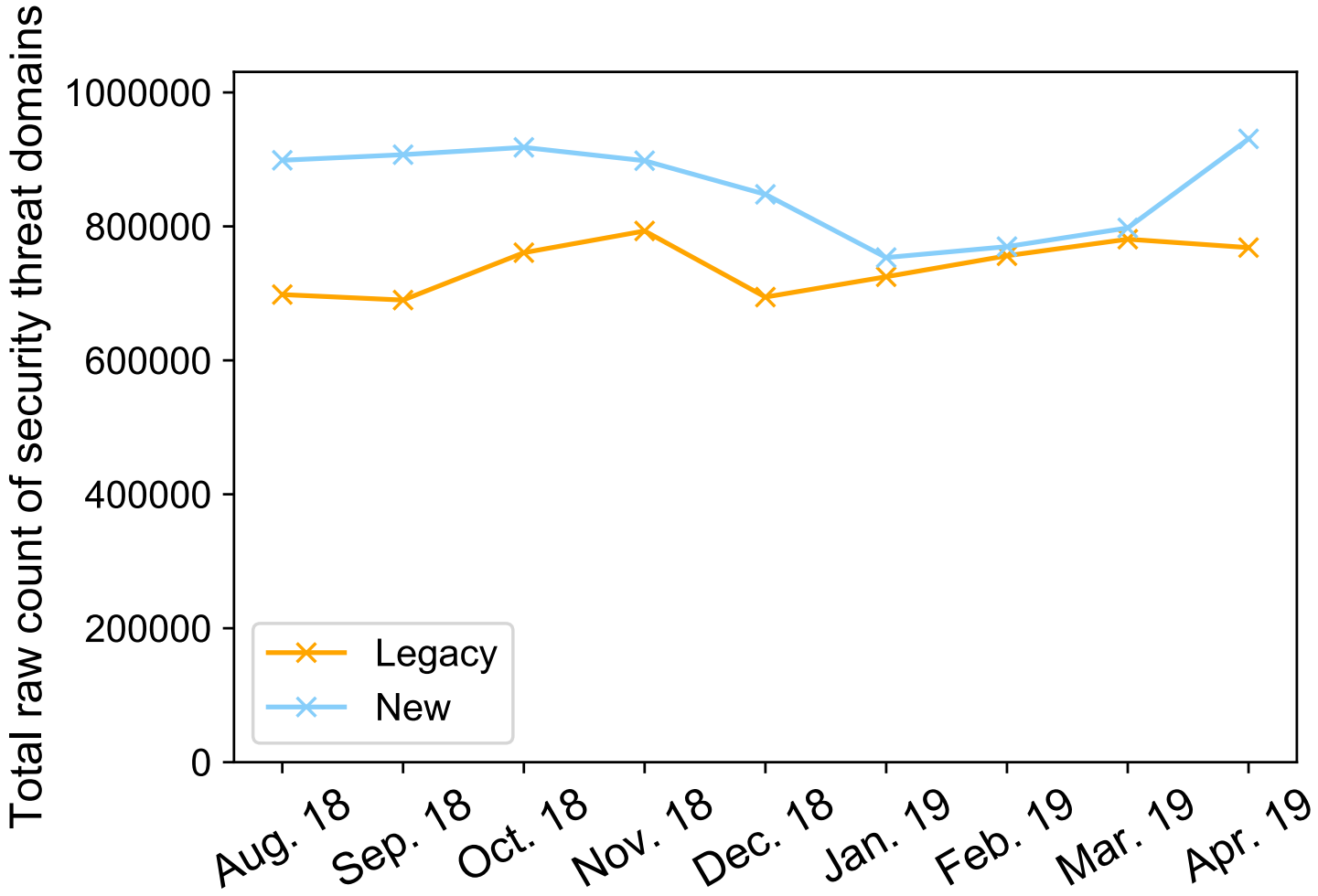
# Is DAAR an Abuse List?

- 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 and is used by
  - Commercial security systems
  - Academia and industry
- These usages show that these datasets exhibit:
  - accuracy, reliability and low false positive rates
  - global coverage

# Abuse Type Distribution



# Number of Domains Identified as Security Threat



# Thank you!

- Root-server analytics
  - Roy.Arends@icann.org
- ITHI
  - Alain.Durand@icann.org
- DAAR
  - Daar@icann.org
  - Samaneh.Tajali@icann.org
  - John.Crain@icann.org

# Engage with ICANN

## Thank You and Questions

Visit us at **icann.org**
Email: email


@icann


linkedin/company/icann


facebook.com/icannorg


slideshare/icannpresentations


youtube.com/icannnews


soundcloud/icann


flickr.com/icann


instagram.com/icannorg