The Tool Formerly Known as Domain Abuse Reporting Tool

Dave Piscitello
on behalf of the ICANN Office
of the CTO SSR Team

A Note from the Lawyers

The ICANN Organization received a Cease and Desist letter for the use of the word “DART”.

We have renamed the project

DOMAIN ABUSE ACTIVITY REPORTING Project (DAAR)

The Domain Abuse Activity Reporting Project (DAAR)

What is the Domain Abuse Activity Reporting Project?

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

How does DAAR differ from other reporting projects?

- Studies all TLD registries and registrars for which we can collect zone and registration data
- Employs a very large set of reputation feeds
- Historical studies
- Studies multiple threats: phishing, botnet, malware, spam
- Scientific approach: unbiased, transparent, reproducible

DAAR Project & the Open Data Initiative

- Goal of Open Data Initiative is to facilitate access to data that ICANN organization or community creates or curates
- DAAR project uses data from public, open, and commercial sources
  - DNS zone data
  - WHOIS data
  - Certain open source reputation data
  - Certain commercial feeds requiring a license or subscription
- In cases where there are no limitations on redistribution of DAAR-related data, these DAAR project data or reports will be published periodically and included in the Open Data Initiative

DAAR Project Goals

Provide ICANN community with data to support the policy development process

- DAAR project data can be used to
  - Identify threats reported at TLD or registrar level for all TLDs for which we can obtain data
  - Historically track security threats, domain registration activity (adds, deletes) at a TLD or registrar level
  - Help operators understand or consider how to manage their reputations, their anti-abuse programs or their terms of service
  - Study malicious registration behaviors
  - Assist the operational security community by sharing open data or data analyzed by the reporting tools

DAAR Uses TLD Zone Data

- Collects zones for TLDs for registry analytics
  - Any {new, legacy, cc} from which we can get a zone
  - Currently gTLDs. Some ccTLD expressed interest in being added during ICANN 58, Copenhagen
- Currently, system collects zones from 1241 TLDs
  - Approximately 195 million domains
  - DAAR project uses publicly available methods to collect zone data (Centralized Zone Data Service, zone transfer)

DAAR Uses Whois

- Collects registration data to associate delegated domain names in zone files with sponsoring registrars
  - DAAR project uses published registration data (Whois)
- DAAR project uses domain names that appear in zones
  - Security threats cannot be executed if a domain name cannot resolve to an IP addresses

DAAR Uses Many Threat Data Sets

- DAAR project 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 billions of users daily
  - Academic and industry use and endorse these data sets
  - Studies and industry use show that they have history of accuracy, global coverage, and low false positive rates

DAAAR project reflects how parties external to ICANN community see the domain ecosystem

- Extensible framework
  - Experimenting with doing analyses using subsets of data

DAAR Is Not An Abuse List Service

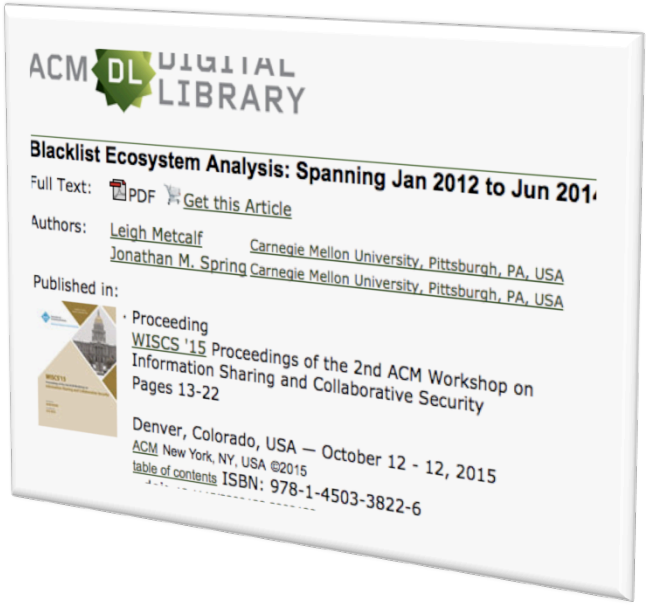
- DAAR project does not identify or investigate abuse
- DAAR project uses multiple domain or URL abuse data sets (reputation feeds) to
  - Counts security threat domain names (spam, phishing, malware hosting, botnet/C2), total abuse domains, cumulative abuse domains
  - Create histograms, charts, days in the life views
  - Search DAAR data by argument
- If a domain appears on any list, it is included in the counts (de-duplication is part of process)

Current Reputation Data Sets

- Spamhaus Domain Block List (DBL)
- Anti-Phishing Working Group eCrime eXchange (eCX).
- Phishtank
- Malware Patrol (composite block list):
  - SpamAssassin: malware URLs list
  - Symantec Web Security
  - Carbon Black Malicious Domains
  - Firekeeper
  - Postfix MTA
  - DansGuardian
  - Squid Web proxy blocklist
  - Ransomware C&C server IPs
  - Smoothwall
  - Mailwasher
  - Symantec Email Security for SMTP
  - Mozilla Firefox Adblock
- Ransomware Tracker: malware C&C servers.
- Feodotracker: domains used to support malware.

Why Multiple Data Sets?

- Expands our abuse data set with low duplication
  Metcalfe & Spring. Blocklist Ecosystem Analysis
  http://dl.acm.org/citation.cfm?id=2808129
- Research finds that there is *little overlap between block lists*
- We use data feeds with
  - Industry reputation for accuracy, clarity of process
  - Threat classification that matches our purposes
  - Consensus adoption across operational security community, i.e., inclusion in commercial security systems
  - Frequency of citation in academic literature



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 % overlap it might have with a larger list

Percentage of Abuse (Experimental)

- Experimenting with measuring abuse impacting TLDs and Registrars
  - Purpose of scoring is to measure the extent to which an operator is a target of malicious actors
- Seeking community input
  - Goal is to gain industry-wide acceptance on scoring

Percent of Abuse, TLD

- The number of unique, currently listed domains per 100 domains in the zone

$$PAB_{ry} = \frac{\text{abuse-listed domains in a TLD on a given day}}{\text{domains in the TLD zone on this day}} \times 100$$

- This shows us the percentage of domains in the zone file that are currently listed on abuse lists that we monitor

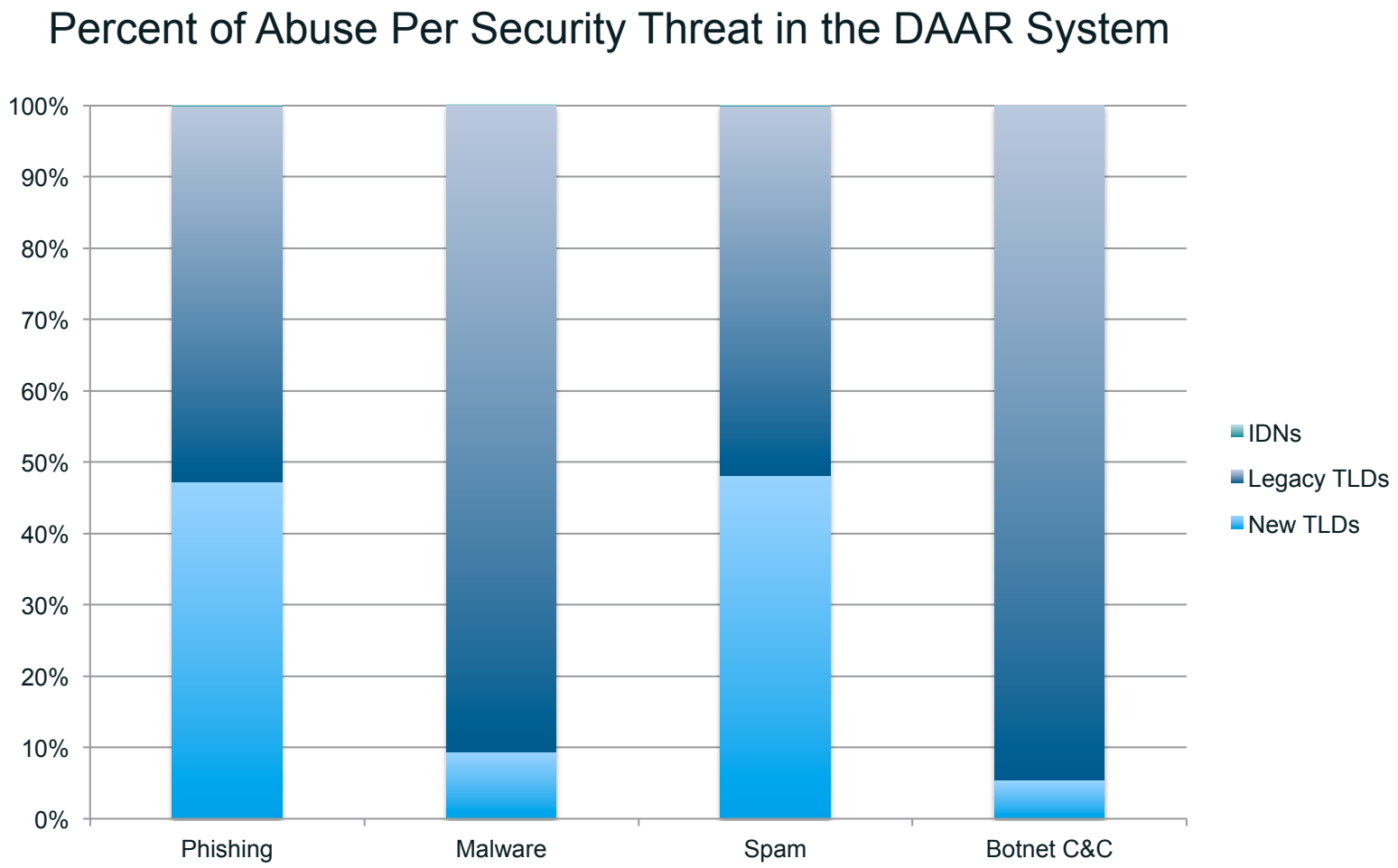
Percent of Abuse, Registrar

- The number of unique, currently listed abuse domains per 100 domains that the registrar sponsors.

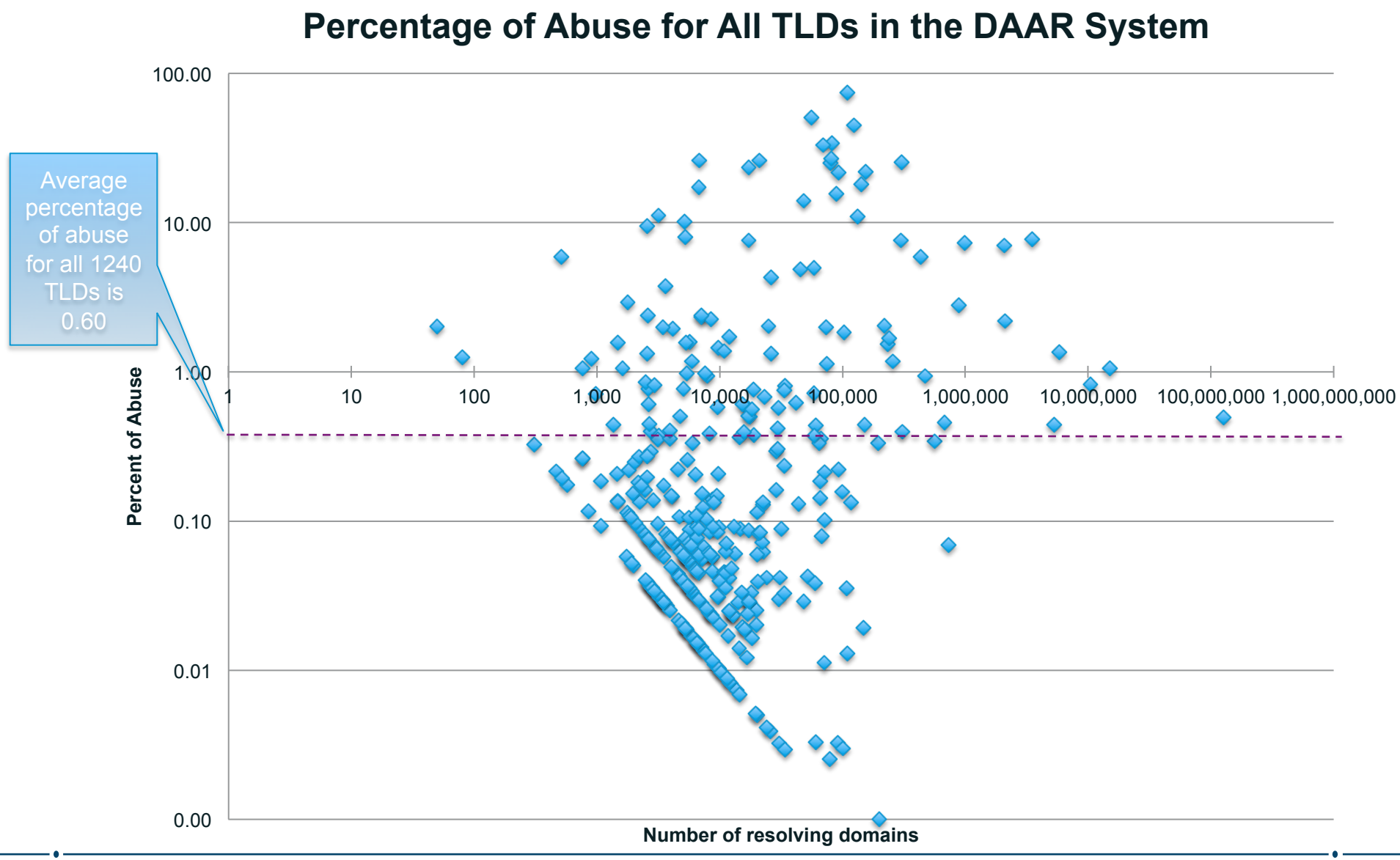
$$PAB_{rr} = \frac{\text{abuse-listed gTLD domains sponsored by registrar on a given day}}{\text{gTLD domains sponsored by the registrar on this day}} \times 100$$

- This shows us the percentage of the domains that the registrar sponsors are currently listed on abuse lists that we monitor.

Sample Statistics



Sample Statistics (Cont'd)



Access to Reporting System

- Currently in Beta, internal use
- Expressions of interest from ccTLD operators
  - Contact the Office of the ICANN CTO to discuss access
- Soliciting community input on kinds and frequency of reporting
  - What should we report?
  - How should we represent data or findings?
  - To whom should we report?
  - Order of reporting?
  - Access to our data?
    (Note: may be affected by use licenses)

Questions?