Domain Abuse Reporting Tool (DART)

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

The DART Project

What is the Domain Abuse Reporting Tool?

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

How does DART differ from other reporting?

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

Goals of the Domain Abuse Reporting Tool

Provide ICANN community with data to support the policy development process

- 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

DART 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 1236 TLDs
  - Approximately 193 million domains
  - Application rejection or renewal issues with ICANN Centralized Zone Data Service

DART Uses Whois

- Collects registration data to associate delegated domain names in zone files with sponsoring registrars
  - Issues with Whois rate limiting
- DART uses domain names that appear in zones
  - Security threats cannot be executed if a domain name cannot resolve to an IP addresses

DART Uses Reputation Data (Blocklists)

- Uses multiple domain or URL abuse data sets (reputation feeds) to
  - Count spam, phishing, malware host, botnet (C2) domain names, total abuse domains, cumulative abuse domains
  - Create histograms, days in the life views…
  - Search abuse database by argument
- If a domain appears on any list, it is included in the counts (de-duplication is part of process)

DART Uses Many Reputation Data Sets

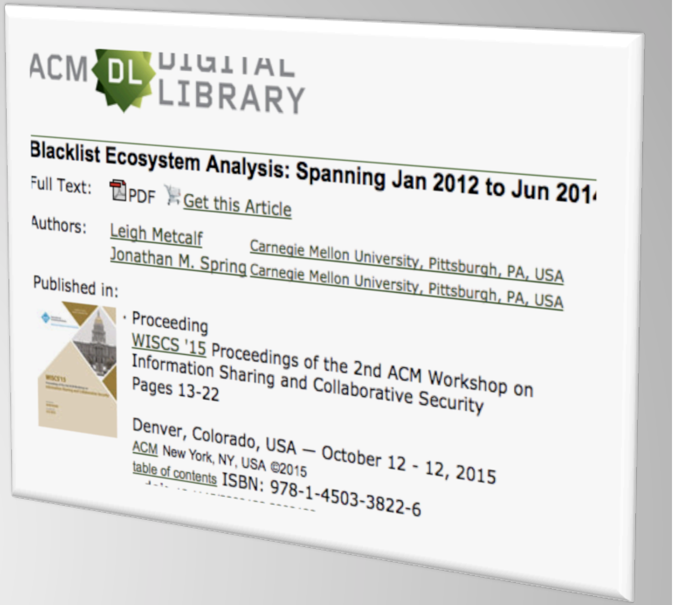
- DART collects the same abuse data that is reported to industry and Internet users
  - The abuse data that DART collects are used by commercial security systems that protect billions of users daily
  - Academic and industry use and endorse these data sets
  - Studies and industry use show that they have history of accuracy, global coverage, and low false positive rates
- *DART reflects how parties external to ICANN community see the domain ecosystem*
- Extensible framework
  - Experimenting with doing analyses using subsets of data

Why Multiple Data Sets?

- Expands our abuse data set with low duplication

  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



ACM DL DIGITAL LIBRARY

Blacklist Ecosystem Analysis: Spanning Jan 2012 to Jun 201

Full Text: PDF Get this Article

Authors: Leigh Metcalf Carnegie Mellon University, Pittsburgh, PA, USA
Jonathan M. Spring Carnegie Mellon University, Pittsburgh, PA, USA

Published in:

- Proceeding
  WISCS '15 Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security
  Pages 13-22

  Denver, Colorado, USA — October 12 - 12, 2015
  ACM New York, NY, USA ©2015
  table of contents ISBN: 978-1-4503-3822-6

More on "Why multiple lists?"

- 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
- Experience with our data sets is similar to Metcalfe & Spring's and Sinha findings

Scoring (Experimental)

- Purpose of scoring is to assess deviation (distance) from mean scoring
  - Measure the extent to which an operator is a target of malicious actors
- Experimenting with strawman proposal for scoring abuse impacting TLDs and Registrars
- Looking for input
  - Goal is to gain industry-wide acceptance on scoring algorithms

Abuse Score, TLD

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

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

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

Abuse Score, Registrar

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

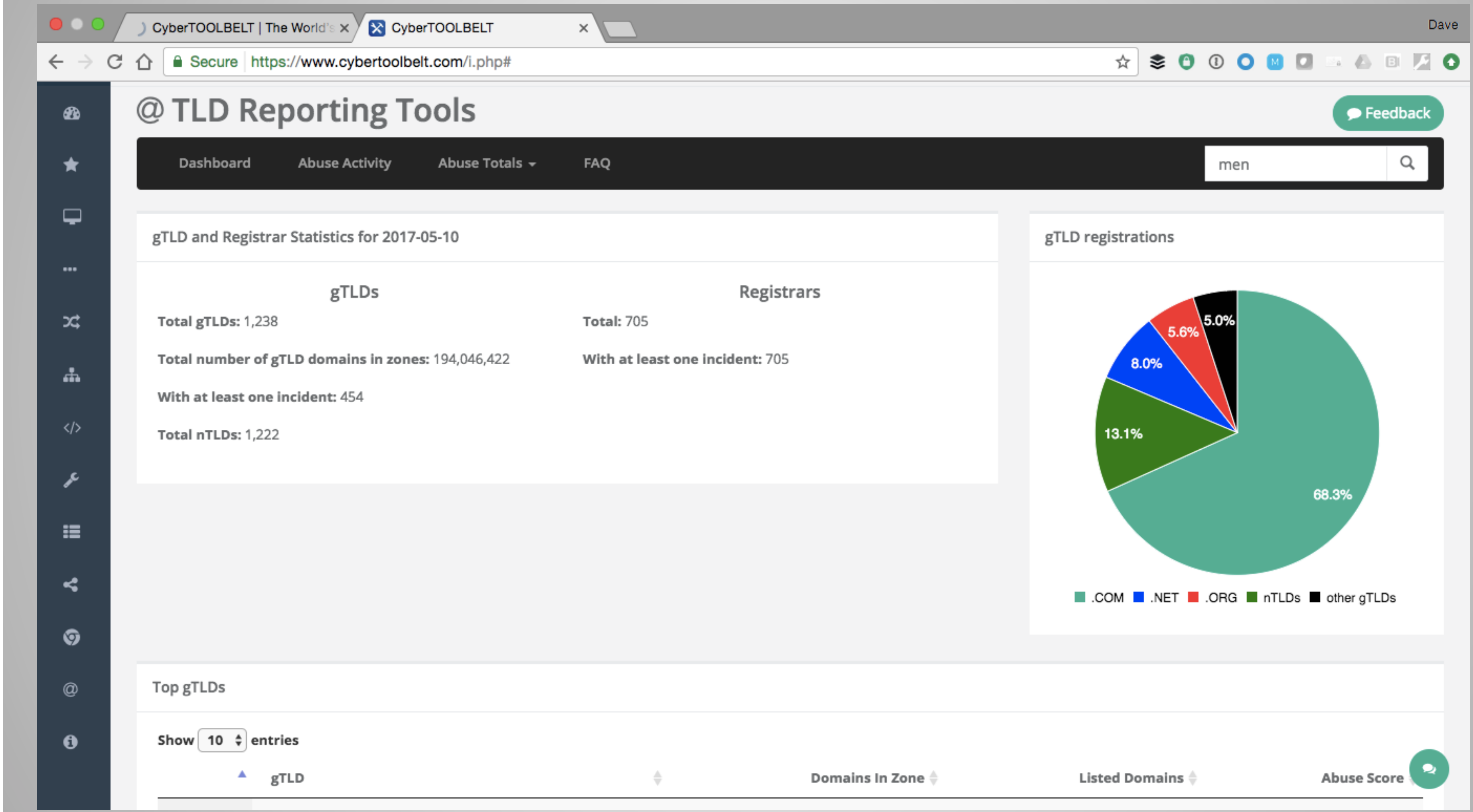
$$\text{SCORE} = \frac{\text{abuse-listed gTLD domains sponsored by registrar on a given day}}{\text{gTLD domains sponsored by the registrar on this day}} \times 100$$

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

Access to Reporting System

- Currently in Beta, internal use
- Soliciting community input on kinds and frequency of reporting.
  - What should we report?
  - To whom should we report?
  - Order of reporting?
  - Access to our data?
    (Note: may be affected by use licenses)

DART Dashboard



@ TLD Reporting Tools

gTLD and Registrar Statistics for 2017-05-10

gTLDs

Total gTLDs: 1,238

Total number of gTLD domains in zones: 194,046,422

With at least one incident: 454

Total nTLDs: 1,222

Registrars

Total: 705

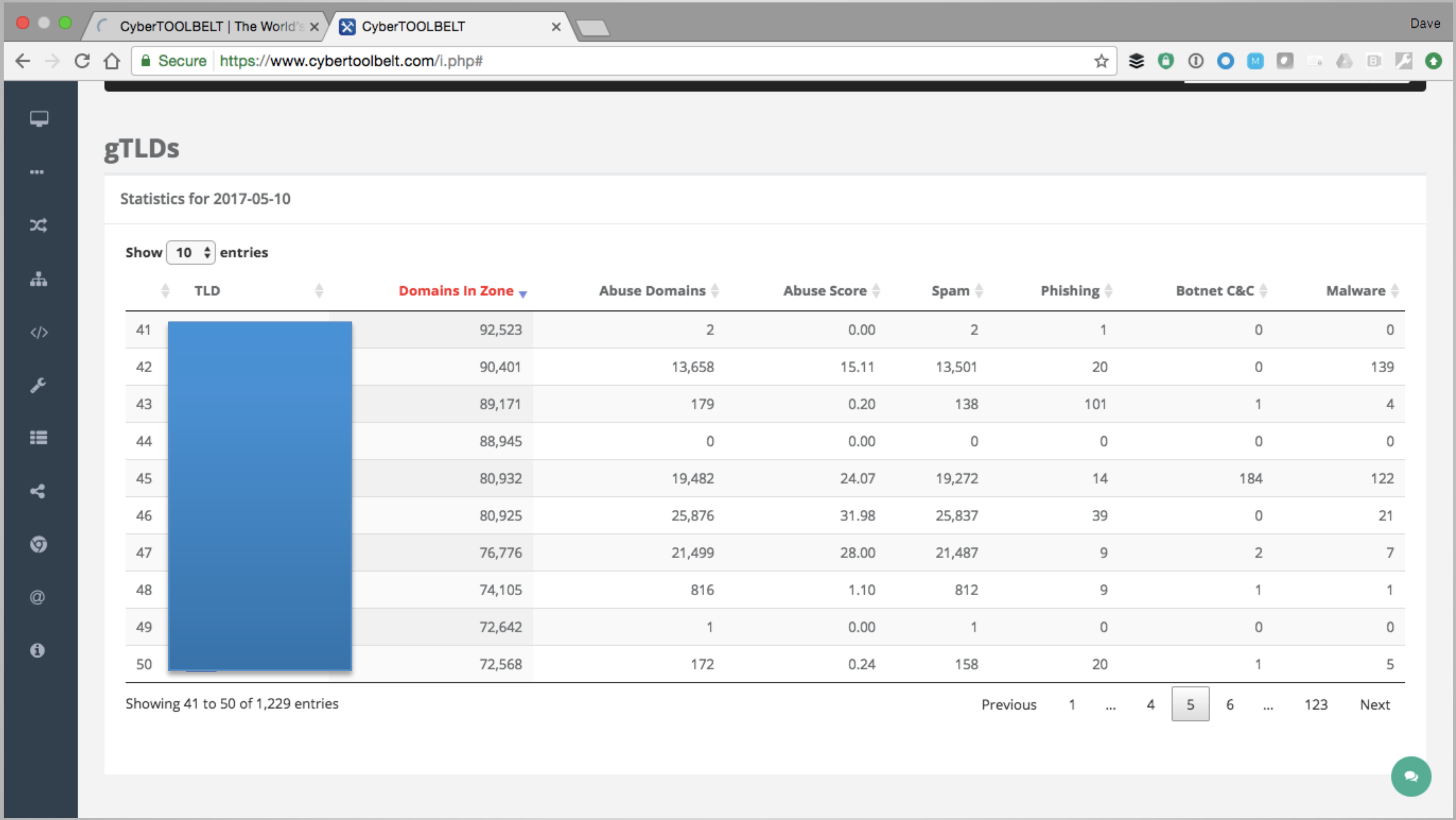
With at least one incident: 705

gTLD registrations

Top gTLDs

gTLD	Domains In Zone	Listed Domains	Abuse Score

Sortable Listings


CyberTOOLBELT | The World's
CyberTOOLBELT
Dave
Secure
https://www.cybertoolbelt.com/i.php#

gTLDs

Statistics for 2017-05-10

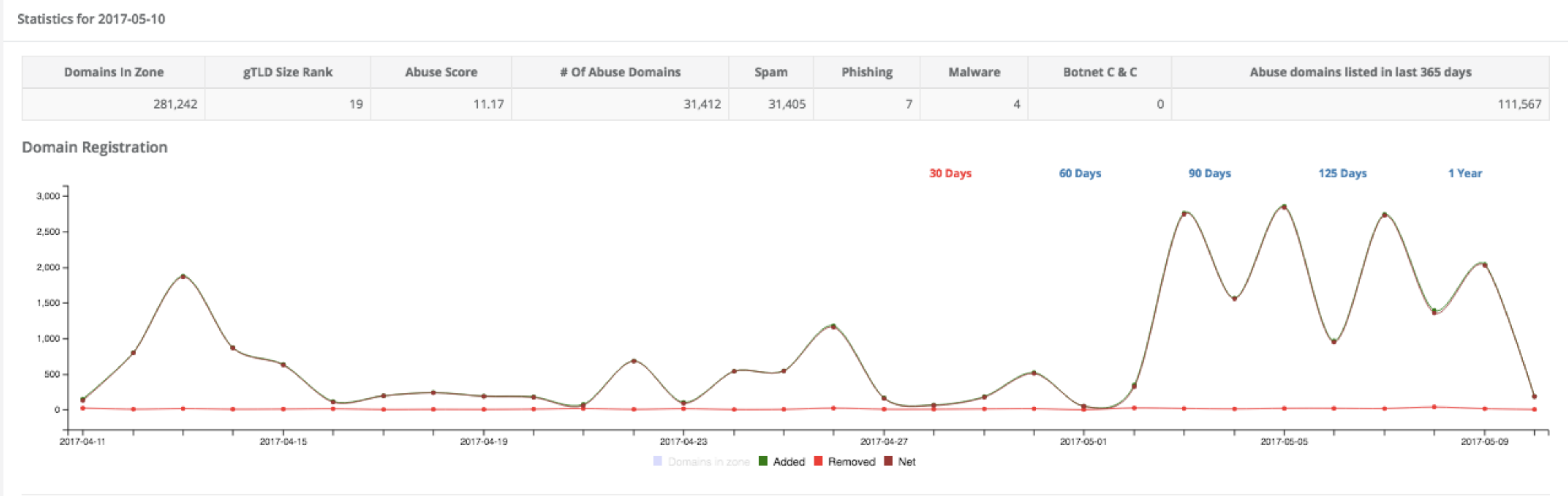
Show 10 entries

	TLD	Domains In Zone	Abuse Domains	Abuse Score	Spam	Phishing	Botnet C&C	Malware
41	[redacted]	92,523	2	0.00	2	1	0	0
42	[redacted]	90,401	13,658	15.11	13,501	20	0	139
43	[redacted]	89,171	179	0.20	138	101	1	4
44	[redacted]	88,945	0	0.00	0	0	0	0
45	[redacted]	80,932	19,482	24.07	19,272	14	184	122
46	[redacted]	80,925	25,876	31.98	25,837	39	0	21
47	[redacted]	76,776	21,499	28.00	21,487	9	2	7
48	[redacted]	74,105	816	1.10	812	9	1	1
49	[redacted]	72,642	1	0.00	1	0	0	0
50	[redacted]	72,568	172	0.24	158	20	1	5

Showing 41 to 50 of 1,229 entries

Previous 1 ... 4 5 6 ... 123 Next

Registration Histories

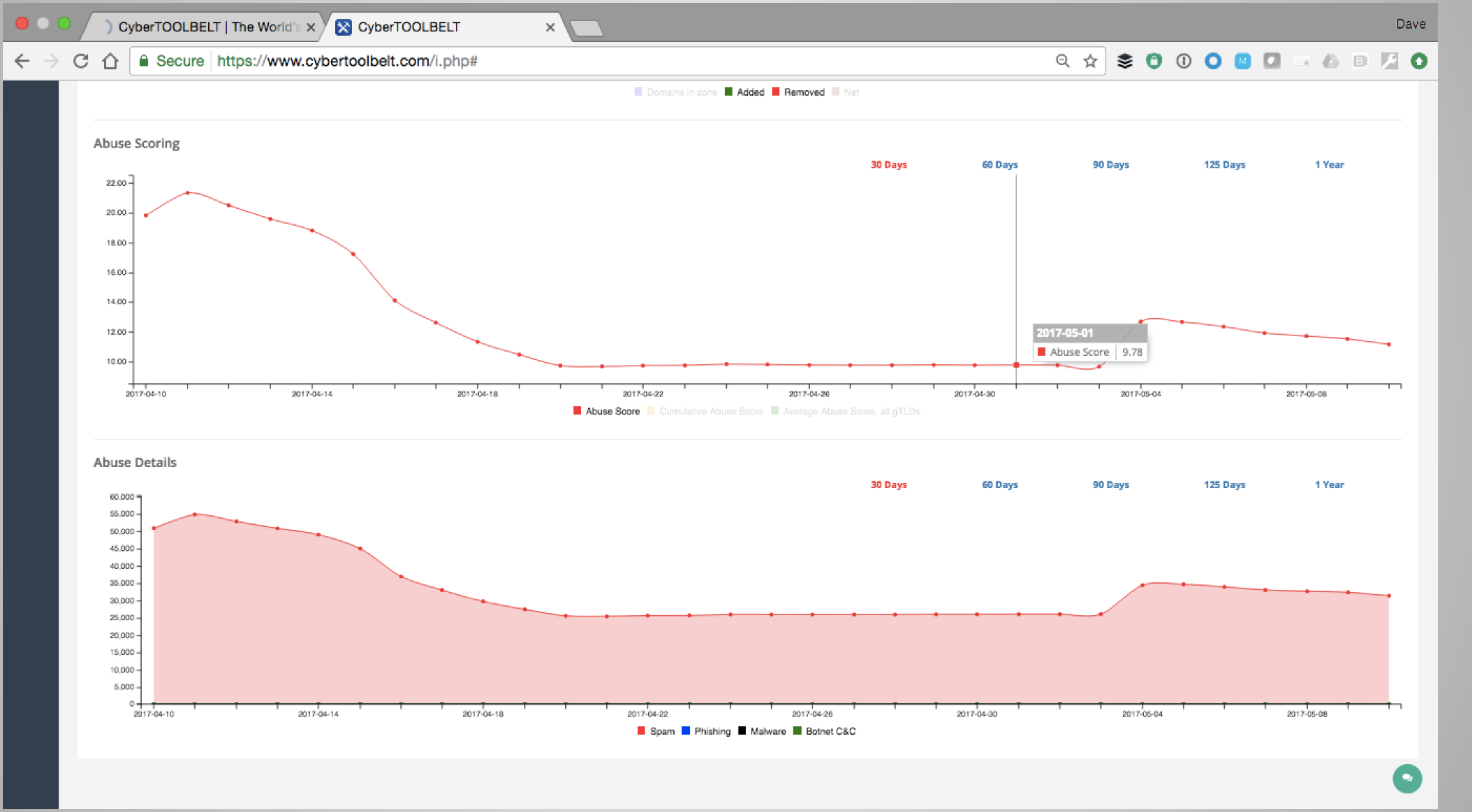


Statistics for 2017-05-10

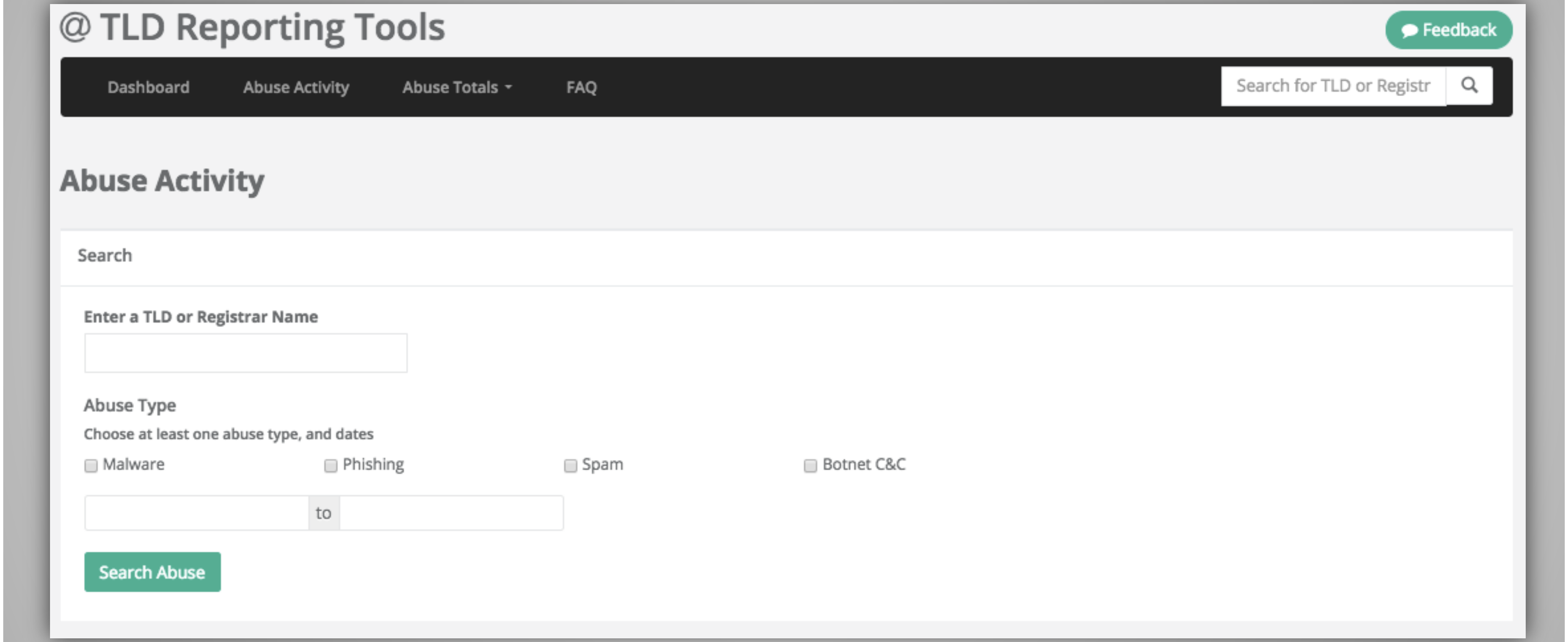
Domains In Zone	gTLD Size Rank	Abuse Score	# Of Abuse Domains	Spam	Phishing	Malware	Botnet C & C	Abuse domains listed in last 365 days
281,242	19	11.17	31,412	31,405	7	4	0	111,567

Domain Registration

Abuse Histories: Details



Search



@ TLD Reporting Tools

Abuse Activity

Search

Enter a TLD or Registrar Name

Abuse Type

Choose at least one abuse type, and dates

□ Malware □ Phishing □ Spam □ Botnet C&C

to