Domain Abuse Activity Reporting (DAAR)

Samaneh Tajalizadehkhooob
Lead SSR Specialist - ICANN OCTO
Outline

• DAAR definition
• DAAR data collection
• DAAR project partners
• DAAR analytics
• Data sharing & API
• DAAR project status
“Systems are particularly prone to failure when the person guarding them is not the person who suffers when they fail.”

Therefore,

Insecurity is as much an incentive problem as it is a technical problem.
Problem

A **growing** need for proactive detection and mitigation strategies by TLD operators & registrars

But there is lack of knowledge about

- Abuse concentrations in TLD networks
- Operators’ abuse performance in comparison to their peers
Domain Abuse Activity Reporting (DAAR)
What is DAAR?

A system for reporting on domain name registration and abuse data across TLD registries and registrars.
What is DAAR?

DAAR data can be used to

• Report on **threat activity** at TLD or registrar level
• Study **malicious registration** behaviors
• 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
Outline

• DAAR definition
• **DAAR data collection & methodology**
• DAAR work breakdown & project partners
• DAAR analytics
• Data sharing & API
• DAAR project status
Data Sources

1. DNS zone data
2. WHOIS
3. Open source or commercial abuse threat or reputation blacklist (RBL) data*

*Certain data feeds require a license or subscription
DNS Zone Data

• Uses
  • Publicly available methods Centralized Zone Data Service (CZDS)
  • Domain names in zone files

• Collects
  • Approximately 1220 gTLDs
  • Approximately 192 million domains
WHOIS

DAAR uses

Published WHOIS registration data

Registrar name and IANA ID

Current challenges

Reliable, accurate registrar reporting depends on WHOIS

Scaling data collection
Abuse Threat Data

DAAR counts “unique” abuse domains

A domain that appears on any abuse datasets reporting to DAAR is included in the counts once
Abuse Threat Data

DAAR uses multiple abuse Reputation Blocklist (RBL) datasets to generate:

- Daily raw counts of domains associated with abuse
- Daily total and cumulative percentage abuse domains
- Calculate monthly/yearly newly added abuse domains
- Visual analytics regarding abuse trends
Data Collection in a Nutshell

1. List of domains in zone
2. DNS Zone Data
3. Blacklist/Blocklist

WHOIS

CZDS

TLD 1

TLD 2

DAAR Reputation Metrics per Registry
Reputation Block Lists: Identifying Threats

DAAR collects domain data for

• Phishing
• Malware
• Spam
• Botnet Command & Control
Current Reputation List

Domains only

- 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)
Outline

• DAAR definition

• DAAR data collection & methodology

• **DAAR work breakdown & project partners**

• DAAR analytics

• Data sharing & API

• DAAR project status
## Work Breakdown

<table>
<thead>
<tr>
<th>Detailed Task</th>
<th>Frequency</th>
<th>Party</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone files</td>
<td>Daily</td>
<td>iThreat Cyber Group</td>
</tr>
<tr>
<td>WHOIS</td>
<td>Daily</td>
<td>iThreat Cyber Group</td>
</tr>
<tr>
<td>Abuse feeds</td>
<td>Daily</td>
<td>iThreat Cyber Group</td>
</tr>
<tr>
<td><strong>Data preprocessing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepossessing all the data feeds to remove anomalies, false positives, and others.</td>
<td>Daily</td>
<td>iThreat Cyber Group</td>
</tr>
<tr>
<td><strong>Data Aggregation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregate all the 3 data types, merge them and produce abuse metrics</td>
<td>Daily/Monthly</td>
<td>iThreat Cyber Group</td>
</tr>
<tr>
<td><strong>Data Analytics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleaning the data and producing aggregated statistics and analytics</td>
<td>Monthly</td>
<td>Samaneh (OCTO-SSR)</td>
</tr>
<tr>
<td><strong>Monthly Reports</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publishing DAAR white paper including monthly &amp; historical analysis of TLD abuse</td>
<td>Monthly</td>
<td>Samaneh (OCTO-SSR)</td>
</tr>
</tbody>
</table>
Outline

• DAAR definition
• DAAR data collection & methodology
• DAAR work breakdown & project partners
• DAAR analytics
• Data sharing & API
• DAAR project status
Abuse Type Distribution

- Spam Domains: 87.8%
- Malware Domains: 3.4%
- Phishing Domains: 6.8%
- Botnet C&C Domains: 2.0%
Distribution of Domains in gTLD Zones

- Legacy: 88.3%
- New: 11.7%
Total Number of Domains Identified as Security Threat

![Graph showing total number of security threat domains from August 18 to April 19. The graph compares Legacy and New categories. Legacy remains relatively stable, while New shows a steady increase.]
How Many gTLDs are Driving the Bulk?

Abused

- 54.8% New
- 45.2% Legacy

Graph showing the percentage of abused domains against the count of new gTLDs.
How Many gTLDs are Driving the Bulk?

Abused
Abuse: raw counts vs normalized counts
Outline

• DAAR definition
• DAAR data collection & methodology
• DAAR work breakdown & project partners
• DAAR analytics
• Data sharing & API
• DAAR project status
DAAR Data Access
(gTLDs only)
What is MoSAPI?

REST API that allows Registries to retrieve information collected by the SLAM.
Getting the latest DAAR data

\[
<\text{base_url}>/\text{daar/report/latest}
\]

curl --cookie cookies.txt
https://mosapi.icann.org/mosapi/v1/example/daar/report/latest

\{
   "version": 1,
   "tld": "example",
   "daarReportDate": "2018-12-12",
   "daarReportData": {
      "domainsInZone": 27957,
      "uniqueAbuseDomains": 14,
      "spamDomains": 10,
      "phishDomains": 3,
      "botnetCcDomains": 0,
      "malwareDomains": 2
   }
\}

Latest DAAR aggregates.
Getting DAAR data

Additional methods to get DAAR data:

• DAAR data for the specified date in the URL
  <base_url>/daar/report/<YYYY>-<MM>-<DD>

• List of dates for which DAAR data is accessible
  <base_url>/daar/reports?startDate=<startDate>&endDate=<endDate>
Getting the latest DAAR data

<base_url>/daar/reports?startDate=<startD\ate>&endDate=<endDate>

curl --cookie cookies.txt
https://mosapi.icann.org/mosapi/v1/example/daar/reports

```json
{
  "version": 1,
  "tld": "example",
  "daarReports": [
    {
      "daarReportDate": "2018-12-12",
      "daarReportGenerationDate": "2018-12-13T23:20:50.52Z"
    },
    {
      "daarReportDate": "2018-12-13",
      "daarReportGenerationDate": "2018-12-13T23:20:51.52Z"
    }
  ]
}
```
For more info contact

Gustavo Lozano – Gustavo.Lozano@icann.org
Outline

• DAAR definition

• DAAR data collection & methodology

• DAAR work breakdown & project partners

• DAAR analytics

• Data sharing & API

• **DAAR project status**
DAAR Progress Timeline

2017
- Published DAAR method paper

2018
- Call for public input
- Started with DAAR monthly reports
- Published SSR responses to comments

2019
- Improving DAAR system and data based on feedback received
- Published API for pushing DAAR data to registries
Project Next Steps

Methodology
- Improving the system based on comments and reviews
- Developing and documenting a process for systematically reviewing feeds

Data
- Adding more feeds
- Discussion about sharing data with registries who are interested in viewing their own data

Results
- Developing similar metrics for registrars and ccTLDs
- Developing new metrics and analytics based on DAAR (e.g., looking at other TLD related attributes in addition to legacy and new)?

Having ongoing discussions with contracted parties and others to keep improving DAAR
Challenges Ahead

- **Registrar level metrics?**
  WHOIS data collection is hard to scale
  Possible solution: daily WHOIS queries only for blacklisted domains or a random sample of domains

- **ccTLD level metrics?**
  Problem: Lack global ccTLD zone file access
  Possible solution: Passive DNS data? Root data?

  This is Work-in-progress ...
Question or Comments?

Contact us:

Daar@icann.org
Samaneh.tajali@icann.org
John.crain@icann.org