The Domain Abuse Activity Reporting (DAAR) System

David Piscitello, ICANN and Greg Aaron, iThreat Cyber Group

Abstract

Efforts to study domain name abuse are common today, but these often have one or more limitations. They may only use a statistically valid sampling of the domain name space or use limited sets of domain name abuse (reputation) data. They may concentrate only on a particular security threat, e.g., spam. Notably, few studies collect and retain data over a sufficiently long timeframe to provide opportunities for historical analysis.

This paper describes our efforts to collect a large body of domain name registration data, and to complement these data with a large set of reputation data feeds that report and classify multiple security threats. We intend that the resulting data repository will serve as a platform for daily or historical analysis or reporting. Ideally, any findings or reporting derived from our system can be independently validated. We thus use publicly or commercially available data so that the reports or findings from any studies that use our system would be reproducible by any party who collects the same data sets and applies the same processing rules. The long-term objective for our project is that this system will serve the community by establishing a persistent, fact-based repository for ongoing collection, analysis, and reporting.

Table of Contents

Introduction and Background

Domain names are critically important resources for most Internet applications. Users rely on domain names to simplify their navigation of the World Wide Web and social media, to identify the sources and destinations of electronic mail or other forms of correspondence, and generally to assist with other activities they conduct on the Internet such as online banking or commerce.

Domain names are equally important resources for malicious actors, who exploit user dependence on domain names when they plan and execute illicit activities. Identity theft, financial fraud, malware delivery, extortion, and the operation of criminal infrastructures are examples of the more commonly encountered abuses of domain names.

Efforts to study domain name abuse are common today, but these often have one or more limitations:

- Few efforts study abuse across all Top-level Domain (TLD) delegations,
- Studies do not typically use large numbers of reputation data sets,
- Studies generally do not assess multiple security threats,
- Few efforts store data over time to provide a basis for conducting historical data analyses,
- Many efforts studying domain name abuse are done in the context of products or services, which can lead to biases on how the security threat-related data are collected and displayed, and

Perhaps most importantly,

- The methodologies and data sources for these studies are often not disclosed, so study results cannot always be reproduced.

These limitations have led some organizations to publish reports on DNS abuse that were subsequently criticized for their questionable accuracy, undocumented methodologies, or non-reproducible findings[1]. After requests from the community, the Office of the Chief Technology Officer (OCTO) concluded that the ICANN community would benefit from having a persistent and reproducible set of data from which domain name or registration abuse analyses could be performed, and set out to meet the community's request.

We have initiated a project to collect a large body of domain name data and have complemented this with a large set of reputation data feeds. Our intention is that the data collected by this system could serve as a platform for analysis. We elected to use data that are publicly or commercially available so that the reports or findings from any studies that use our system would be reproducible or independently

[1] E.g., The Web's Shadiest Neighborhoods, and a rebuttal, Domain Dunce Award.

validated by any party who collects the same data sets and applies the same processing rules to those data.

An important concept and goal for the DAAR project is to show how the world outside the ICANN community perceives abuse in the domain name space. DAAR data shows us what domain names are being blocklisted. Reporting data from this perspective benefits us in two ways: (1) it identifies for us the set(s) of domain names that professional anti-abuse sources are concerned about, and (2), it tells us what domains are being blocked by security administrators at private and public networks and services around the world. This is a practical measurement of what is happening in the real world. DAAR data are also especially useful to discover abuse domain registration and usage trends, and to isolate possibly large concentrations of problems in certain parts of (or delegations in) the namespace.

DAAR does not attempt to measure mitigation activity, i.e. it is not intended to measure how various parties (including registries and registrars) respond to abuse activity. DAAR also does not attempt to independently validate or measure whether any particular threat – for example, a malware uploading site – is active at a given point in time. Such measurements are exceeding complex to perform at scale, and are beyond the current scope of the project.

Purposes of the DAAR System

The overarching purpose of the DAAR system is to give the ICANN community reliable, persistent data that can be used to make informed decisions. Within this broad framework, the DAAR system can serve several specific purposes:

- Facilitate registry, registrar, academic or industry studies of malicious domain name registrations, and assist in anti-abuse investigations;
- Provide a means to determine and report on the presence or prevalence of security threats at a gTLD registry or accredited registrar level;
- Track market activity such as domain registrations (adds, deletes) over time;
- Isolate or assist in the identification of the causes of abnormal registration activities;
- Support the ICANN community's consumer confidence and trust activities; and
- Assist ICANN's Contractual Compliance department should it ask for additional information relating to a complaint filed against an accredited registrar or gTLD registry operator.

DAAR System Overview

The DAAR system was designed by members of the OCTO team in collaboration with iThreat Cyber Group (ICG), which is contracted to develop the DAAR system for the

ICANN Organization. The DAAR currently has two major components: a collection system and a reporting system.

The Collection System

The **collection system** currently gathers the zone files of every gTLD[2]. The collection system constructs and maintains a persistent database of the set of domain names in each TLD that can resolve in the DNS.

The collection system then queries TLD registry Whois servers to obtain domain name registration data (Whois) for those domain names. These data provide the ability to study or perform analysis on a per-registrar basis. DAAR uses non-personal identifying data to associate a domain name with the sponsoring registrar.

The collection system also gathers domain abuse data from multiple, independent, security threat-reporting sources (e.g., reputation block list, or RBLs). The reputation feed providers continually add and remove domains and URLs from these lists, according to their own criteria. The collection system collects these updated data from each provider several times per day and de-duplicates domain names as part of processing.

DAAR System Data Collection

Gathering all the necessary data is a large task. The goal of collecting and warehousing data for all delegated gTLDs, which currently approaches 200 million domains, requires that we collect and store gTLD zone information daily. We also collect each reputation feed multiple times per day to monitor for changes, as the reputation provider allows. We query TLD registry Whois services to collect sponsoring registrar information needs and as constraints permit.

Data collection activity of this scale and diversity faces several challenges.

Top-Level Domain Zone Data

To our knowledge, registry operators do not distribute complete and up-to-date lists of all of the domains in their registries to anyone (ICANN organization or any other party). However, gTLD registries are required to make their zone files available on a daily basis. The zone files provide a list of the domains in the TLD that can resolve, i.e. function on the Internet. This list provides a consistent and reliable daily count of the resolving domains in any gTLD registry, and is a reasonable approximation for the number of domains that a registry or registrar has under management.[3]

---

[2] Several country code TLD operators have expressed interest in participating in the DAAR system. We are engaged in discussions with these operators to arrange for daily collection of zone data.

[3] A registry usually contains some domain names that are registered but do not appear in the zone file. These typically include a small number of domain names that do not have delegated

We access zone files for all gTLDs daily using publicly available methods. DAAR collects most zone files via the Centralized Zone Data Services (CZDS). Those that are not available via CZDS are obtained using the gTLD operator's documented access method.

Reliable zone file availability is critical for this project, so we must manage circumstances where zone files are unobtainable[4]. For our purposes, a zone file is unobtainable when:

a) The CZDS service experiences an outage or other technical problem, or

b) Access permissions for an individual registry's zone file expires, or

c) A registry operator denies our access to its zone file, or

d) A registry fails to upload a valid zone file, or

e) Some other technical issue occurs, such as a failed zone file transfer.

If a zone file is unobtainable, and depending on the nature of the disruption, DAAR will not have registration counts for the duration of the outage. This affects reporting in the following ways:

- The DAAR system will not plot a point in any graph (e.g., domains in zone) that makes use of registration count for the day(s) involved.
- The DAAR system will not calculate a new percentage of abuse for that day(s).
- The DAAR system will not plot a point in any graph that reports percentage of abuse for the day(s)
- Summary reports for the date(s) where the DAAR system has not acquired data will contain the (last) available percent of abuse.

The DAAR system does not make the presence of a domain in the zone a pre-condition of being listed as an abuse domain. The DAAR system reflects whether a

---

nameservers, reserved domain names that are not resolving, and domain names that are in the Redemption Grace Period (RGP). (Per ICANN policy, domain names in the RGP are expired and at the ends of their lifecycles, are removed from the zone file, and unless redeemed are then purged from the registry. At a given time, the domain names in RGP may represent two or three percent of the domains in a registry, or higher depending upon renewal rates.) While it is possible to count names in RGP toward the "domain names under management" in a TLD or register portfolio, this requires making a large number of extra WHOIS queries and is subject to some frictional inaccuracy. It is never possible to know with perfect accuracy how many undelegated domain names and reserved domain names are in a registry. In contrast, the number of domain names in the zone file can be counted reliably every day, and can be confirmed by other zone file subscribers.

[4] The SSAC's "SAC097: Advisory Regarding the Centralized Zone Data Service (CZDS) and Registry Operator Monthly Activity Reports" (12 June 2017) notes that "[CZDS] Policy and process difficulties prevent subscribers from gaining and then maintaining reliable access to zone files. These problems affect the ability of subscribers to perform research and security functions that benefit the public interest."

domain name is on a blocklist, not whether it in the zone *and* on a blocklist. This method accurately reflects what is on the blocklists, and therefore how the blocklist providers perceive threats.[5] Also, while changes to blocklists are available on a continual basis, fresh zone files are available only once per day. That means that changes to a zone file can only be determined once every 24 hours, while some abusive domains resolve for only a short time.

DAAR does not attempt to independently validate or measure whether any particular threat – say, a phishing site—is active or offline at any given point in time. Such measurements are very complex to perform accurately and at scale, and are beyond the current scope of the project. Instead, DAAR reflects how long the threat feed providers decide to list a given domain. Each has its own criteria[6].

At this time DAAR does not attempt to measure *mitigation* activity, which is whether a given security threat is addressed or dealt with. That is an entirely separate endeavor that has many dimensions and is currently out of scope of the project.

Sponsoring Registrar Registration Data

The DAAR system uses the publicly available Whois service from registry Whois servers to obtain the sponsoring registrar of each domain name in each TLD zone file. The DAAR system must perform a successful Whois query to obtain the domain name registration data for every domain. The system may query a domain more than once a year, in order to track the domain through its lifecycle and to determine when a domain has been purged from a registry and no longer exists. The total number of queries that the DAAR system must perform is therefore in the tens of millions per month, and hundreds of millions per year.

A Whois query is the only means available to obtain the identity of a domain name's sponsoring registrar: indeed, providing this data is one of the primary purposes of a registration data system. Whois is the common public means that anyone attempting to reproduce the DAAR system's results would use. As part of this project, we have encountered Whois rate limiting by some registries. Rate limiting sometimes makes gathering the data very difficult. It impedes our ability to keep up with daily changes to some zone files, to attribute domains to specific registrars, and to track each domain's registration status as it proceeds through its life cycle from creation to potential expiration and purge from its registry.

---

[5] Some security threats are mitigated by removing the domain from the zone, while others are not. Some abuse domains are deleted and then re-added to a zone over time, in some cases by the same abuse actor. In this paper, we refer to this behavior as recidivism.

[6] RBL providers use different criteria to determine when to add or remove a domain name associated with abuse from their lists. Certain RBL providers consider the risk of recidivism, association of a domain name with a Domain Name Algorithm (DGA), registrant, name server, or other criteria before removing a domain from an RBL. Thus some domain names remain on block lists even after the name has been removed from the gTLD zone file.

Domain Reputation Data (Security Threat or Abuse Data)

The DAAR system integrates data from approximately twenty (20) reputation feeds. The feeds satisfy the criteria we set for the project:

- Reputation in the operational security community and academia for accuracy and a very low false-positive rate (See Annex B, *Academic Studies or Research Involving Blocklisting*).
- Practical, widespread adoption by large numbers of users among industry, governments, academia, etc. Practical adoption serves as an indicator of confidence: these RBL data are actively used by many kinds of commercial and non-commercial entities worldwide to protect their networks and their users.
- Good practices for maintaining the lists, including public procedures for getting domains de-listed.
- High availability (uptime).
- Size and quality of detection infrastructure. The DAAR system requires global coverage. By employing many RBLs, we are able to expand coverage to the union the set of all of the RBL infrastructures.
- Use of classifications or sub-classifications that allow us to place domains into the security threat categories that we study: phishing domains, malware domains, botnet command-control domains, and spam domains.

The DAAR system obtains updated data from its reputation sources multiple times per day, in certain cases as frequently as two hours. To date, the failure of one download attempt has rarely affected daily counts in the DAAR system.

If the DAAR system fails to receive data from an abuse list, then any domains on that blocklist continue to be counted as abuse domains until we receive the next update of that blocklist (and no newly blocklisted domains are added). That next update of the blocklist may contains some adds and some deletes, and we account for those accordingly. Since the system retrieves data from the blocklists multiple times a day, a temporary service interruption rarely affects DAAR's daily statistics.

DAAR Reporting System

The second component of the DAAR system, the **reporting system**, calculates domain registration and security threat measurements (e.g., counts and percentages) using data stored by the collection system. The DAAR system currently maintains the following data or measurements:

- Domain names contained in the zone files (per registry, registrar).
- Domains that are listed as abusive in at least one reputation data feed that the DAAR system tracks (per registry, registrar). This is a total of unique domains; see the section entitled *DAAR Threat Data Computation and Curation*.

- Percentage of Abuse. This is a measure of abuse reported to the DAAR system by the reputation data feeds that the DAAR system employs compared against the total number of domains in the TLD zone file (per registry, registrar). See the section entitled *Percentage of Abuse*.
- Individual measures of the number of abuse domains that have been classified as phishing, botnet command and control (C&C), malware, and spam domains, using the reputation data feeds that the DAAR system employs (per registry, registrar), see the section entitled *Security Threats observed by the DAAR System*.
- Abuse domains listed in the last 365 days. This is a cumulative count of the domains that have been added to at least one reputation data feed that the DAAR system employs (per registry, registrar).
- Newly registered abuse domains. This is a monthly cumulative count of the domains that have been added to at least one reputation data feed that the DAAR system employs (per registry, registrar).

ICANN staff can export these data or measurements for report generation purposes. The DAAR administrative interface allows ICANN staff to view summary and individual operator measurements in tabular formats. The reporting system plots historical data and automatically generates a pre-defined set of charts. The reporting system also has search and export data functions[7] that ICANN staff use to conduct additional open source intelligence (OSINT) operations on domain names that appear in search results (e.g., DNS or Internet addressing analysis).

Security Threats Observed by the DAAR System

The DAAR system identifies and tracks domain names associated with four security threats (abuse types):

- *Phishing domains* – Domain names that support web pages that masquerade as a trustworthy entity such as a bank or online merchant. Phishing is often associated with financial fraud but is also used to steal identities, domain registration accounts, personal email or email contact lists, and more.
- *Malware domains* – Domain names that facilitate the hosting and/or spreading of hostile or intrusive software that is installed, potentially

[7] The current licensing agreements that we have with certain commercial blocklisting services limits our use of certain data exported though the search function; specifically, we are not permitted to share or publish lists of specific domain names. Should such lists be of value to the ICANN community, we can revisit licensing with the providers.

without the permission of the user[8]. Malware infection counts often include Trojan software, rootkits, ransomware, their variants or families.

- *Botnet command-and-control domains* – Domain names that are used to identify hosts that control botnets. Botnets are collections of malware-infected computers that can be used to perpetrate various abusive activities[9].

These three security threats were specifically listed among others by the ICANN Government Advisory Committee (GAC) Beijing Communiqué of 11 April 2013, which led to a requirement in the new gTLD contracts:

> *"Registry Operator will periodically conduct a technical analysis to assess whether domains in the TLD are being used to perpetrate security threats, such as pharming[10], phishing, malware, and botnets. Registry Operator will maintain statistical reports on the number of security threats identified and the actions taken as a result of the periodic security checks." [Specification 11, paragraph 3b]*

DAAR also collects a fourth category of abuse, *spam domains*:

Spam domains – Predominantly domains that are advertised in the body of bulk, unsolicited email messages, and also domains used to support a spam delivery infrastructure that can be used to distribute the other security threats. A percentage are domains spammed in other ways, such as via SMS. Domain names that are counted as spam domains are primarily domain names that are extracted from URIs that are found in email message bodies or attachments (for example, Adobe PDF or Flash, MS Office documents, or scripting language files) to identify harmful sites or content that is malicious or fraudulent. Domain names that are encountered while analyzing spam delivery infrastructures are also counted because spam delivery infrastructures are particularly exploitative of Internet identifiers, particularly domain names and Universal Resource Identifiers (URIs). In cases where the blocklist operator is able to determine that the sender domain in an email message is malicious, those domain names are counted as spam domains as well. Blocklist operators may also include domain names that are extracted from URLs in text, SMS (cellular carrier submissions), comment or other forms of messaging spam.

---

[8] AV-Test Institute claims to register 390,000 new malicious programs every day and publishes charts that illustrate total malware over time.

[9] The Shadowserver Foundation maintains statistics, maps, and charts of botnet activity and malware evolution.

[10] In the article, Lending Clarity to Security Risk Definitions for ICANN Community and Beyond, the authors observe that "Pharming is not as easily detected as phishing and it hasn't been as well documented. Since pharming changes the translation of domain names to servers controlled by the attacker, there are no URL blocklists dedicated to pharming." To date, we have been unable to identify or collect sufficient or reliable pharming data.

Why Track Spam Domains?

The ICANN Governmental Advisory Committee (GAC) expressed interest in spam as a security threat in its Hyderabad correspondence[11] to the ICANN Board of Directors.

Further, spam is a major means of delivery for the above-listed security threats. Second, most spam messages are sent via illegal or duplicitous means.

Many spam messages are sent via methods that are criminal. The majority of the spam messages sent worldwide are sent via botnets, which are networks of hijacked computers infected with malware. Some spam campaigns use hijacked IP space, e.g., IP addresses that attackers steal from legitimate address registrants or use without permission. *Snowshoe spammers* use many IP addresses or domain names so that they can (temporarily) evade antispam measures. Spammers regularly use many fictitious business names, fake names and identities, and other methods to hide their identities, often in violation of national laws. In contrast, legitimate mailers try hard to build a reputation based on opt-in subscription, a real business address, a known domain, and a small, permanent, well-identified range of sending IPs.

Spam is no longer singularly associated with email. Spammers have adapted to the emergence of social media and other forms of correspondence or merchandising. Today, link spam[12], spamdexing,[13], tweet spam and text message spam pose security threats to our increasingly mobile or social media-focused societies. DAAR includes spam as both a security threat and an indicator of abuse in the domain name system.

Many spam messages are sent via criminal infrastructures that serve as a delivery infrastructure for many forms of criminal activity, especially financial fraud (phishing, ransomware) and malware distribution. The Avalanche network is a recent and noteworthy example of how spam has evolved from a word that identifies nuisance or unsolicited messages to term applied to criminal network or service infrastructures with cloud computing characteristics. Europol hosts an informative infographic that describes how the Avalanche network exploits Internet identifiers in some detail. A copy of this infographic is attached as Annex D.

Sources vary, but around 75% of all email message sent worldwide are spam. The Cisco 2017 Annual Cybersecurity Report states that, "spam accounts for nearly two-thirds (65 percent) of total email volume, and our research suggests that global spam volume is growing due to large and thriving spam-sending botnets… about 8 percent to 10 percent of the global spam observed in 2016 could be classified as malicious

---

[11] See "Questions to the ICANN Board on DNS Abuse Mitigation by ICANN and Contracted Parties"

[12] Also known as comment spam, link spam refers to the insertion of out-of-context hyperlinks on social media or websites to increase the number of external links to a site the spammer promotes.

[13] Also known as SEO spam, spamdexing refers to creation of websites with content that is crafted to cause the website to be indexed with a high position in the search engines.

The DAAR system counts spam domains, not spam messages. Spam domains are one of the most visible measures of reputation to industry and Internet users. Spam domain reputation influences how extensively or aggressively security or email administrators apply filtering for their organizations. Specifically, spam domain reputation can affect whether an organization decides to filter based on individual addresses or names, or on coarser levels such as autonomous system or top-level domain.

DAAR Threat Data Compilation and Curation

DAAR system uses security threat data that are processed and reported by open or commercial reputation data providers. It provides information about the actual domain blocking and abuse identification that is used by organizations of diverse kinds worldwide to assess, mitigate, or eliminate security threats.

The DAAR system collects and compiles domain name reputation data from multiple, publicly available sources (feeds) to identify and classifies the data into the four types of security threats mentioned above. The reputation feed providers meet the criteria we set for this project: accuracy, coverage, industry adoption, ability to redress false positive events, and the feed's ability to classify events into the security threat classes that DAAR tracks.

We use the term *curation* to describe the processes necessary to extract domain names from the host name or URL formats we encounter in the reputation feeds. De-duplication is an important part of the curation process. A domain name that we extract from any reputation data list will cause the domain name to be counted *once* as an abuse domain when we calculate *unique* abuse totals[14]. If a domain is listed for two or more types of abuse, that domain will be counted (again, once) in each relevant abuse category. For example, if one blocklist identifies a domain name as a spam domain and a second identifies that domain as a phishing domain, *and* that domain was not listed before, then:

- The total unique abuse count is increased by one,
- The cumulative abuse count (abuse over past 365 days) is increased by one,
- The spam domain count is increased by one, and
- The phishing domain count is increased by one.

The DAAR system looks at the blocklists, determines what TLD each listed domain name is in, and assigns it to its parent TLD accordingly. The DAAR system will count second level domains in registries that allow registration of second-level domains, and third-level domains in the TLD registries that offer third-level registrations. The latter is mainly relevant to ccTLDs: example.co.uk or

[14] We currently report daily abuse totals and a cumulative total over the past 365 days.

example.net.nz is a domain name for our purposes and would be counted in zone file counts, and it will be classified as abuse if it is added to a blocklist. We do not count example.uk.com, foo.uk.com, dave.uk.com, and uk.com as four domain names: we count it as one (uk.com only).

Certain blocklists list subdomains (third-level domains that are not in TLD registries, like example.uk.com) because there are subdomain providers that offer third-level domains (afraid.org, dyn, etc.) and some of those domains have been identified as hosting a security threat (e.g., malware or phishing). Blocklist providers list the third-level domains as appropriate, thereby limited blocking, and avoid listing just the second-level domain so all services on that domain are not blocked by their clients. The DAAR system reduces cases like this to a single domain, e.g., uk.com, which is the domain name that appears in the TLD registry.

The DAAR system makes use of URI blocklists but only counts a domain name once irrespective of how many URLs have been reported that contain that domain name: example.com would be counted once when the DAAR system process URLs such as http://example.com/bank1/login.html, http://example.com/bank2/login.html… http://example.com/bank100/login.html. The DAAR system does not currently report such *URL amplification* but has the data to do so should this activity become relevant to the project or be of interest for an academic study and paper. The DAAR system could also study subdomain delegation if this were relevant.

Reputation Data

Currently, the DAAR system incorporates data from the following reputation data providers:

- SURBL maintains lists of domain names found in spam, domain names used for phishing, and domains used to support malware. We use all of the SURBL sub-lists and tagging. SURBL analyzes and presents original data. It also incorporates listings from other blocklists including **SpamCop, AbuseButler, Malware Domain List, and PhishLabs**. We currently do not use SURBL IP blocklists.
- Spamhaus Domain Block List (DBL) maintains lists of domain names advertised in spam; domains used for phishing; domains used to support malware; and a botnet Command and Control (C&C) domain list. We currently do not use Spamhaus IP blocklists.
- Anti-Phishing Working Group. We curate and extract domain names from the APWG's phishing URL blocklist feed. We count only verified phishing URLs.
- Phishtank. We curate and extract domain names of the verified phishing URLs from this URL blocklist feed.
- Malware Patrol reports domain names used to support malware. This vendor sources and analyzes original data about malware domains. Malware Patrol's

feed also incorporates listings from several other malicious domain blocklists. The DAAR system uses the following sources within Malware Patrol's composite service:

- Malware Patrol domain list (Note: We do not incorporate Malware Patrol's list of domain generation algorithms, DGAs.)
- 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
- Ransomware Tracker reports domain names that host malware command and control (C&C) servers.

Collectively, these providers give multiple sources of abuse listings for the security threats that the DAAR system can measure or report. We employ at least two sets of reputation data for each security threat that the DAAR system tracks. The collection system is pliable and extensible: in the future, we may add or remove blocklists to ensure quality data or to experiment with reporting from reputation data subsets.

The RBLs that the DAAR system uses are used nearly ubiquitously by public and private organizations across the Internet today to protect their users from various kinds of threats. For example, they facilitate the blocking of dangerous domain names or malicious download URLs in web browsers and harmful URLs in spam email filtering. The article, Reputation Block Lists: Protecting Users Everywhere, describes the many adoptions and application of RBLs that make this security service arguably ubiquitous.

Selection of Reputation Data

Each of the reputation data feeds that the DAAR system uses has a positive track record within the operational security community, and that reputation has been maintained over time. Large numbers of organizations rely upon the same feeds that the DAAR system employs to protect their users. Spamhaus, for example, has been operating for 18 years and their lists protect over 3 billion mailboxes. Referrals or queries to the same block lists that the DAAR system uses are incorporated into organizational or ISP-class commercial and open source firewalls, anti-spam gateways, antivirus software, intrusion detection, application content filtering, web proxies, or email server anti-spam systems that protect billions of users daily against security threats.

The operators of the reputation data feeds that DAAR employs each have a well-defined process for adding and removing domains from its lists (for example, see Spamhaus or SURBL). Some domains that appear on blocklists are compromised domains, i.e., domains that have been registered for legitimate use but an attacker has succeeded in gaining administrative control of either DNS or content hosting associated with these domains. We have made concerted efforts to only employ reputation feeds that ensure that most of the domains on the blocklists currently used for this project were registered for the purpose of perpetrating abuse[15]. We have selected blocklist operators that have been studied by academia or found to have satisfactorily low false positive rates (see Annex B). A generic description of how Universal Resource Identifiers (URIs, typically domain names or IP addresses) are determined to be malicious is provided in Annex A.

Blocklist operators are constantly subjected to academic and industry scrutiny. This encourages blocklist operators to list domains that have no redeeming uses, and this same scrutiny creates strong, often economic incentives to not block compromised domains being used for legitimate purposes.

Multiple Reputation Data Sources

The DAAR system is intended to provide persistent, reliable data that others can analyze to assess how abuse in the name space affects or shapes the security threat landscape. Such analyses may be useful in policy deliberation. TLD registry operators may use DAAR data or analyses to adjust mitigation strategies.

Our goal is to capture data that depicts how the Internet community sees and assesses the gTLD space, at least for those gTLDs for which we can obtain data. To reach this goal, we use a very large set of data from multiple reputation feeds. Each of these is assembled differently, so our data set provides a more comprehensive look at the namespace, and data about different types of abuse than studies that use one or few. We reviewed relevant research to learn whether using many feeds would create unacceptably high duplication results. Metcalf and Spring (see Annex B) concluded from their research that the overlap among blocklists is relatively small. We observe a larger overlap than Metcalf and Spring but the duplication does not tax the system and for our purposes, duplication serves as confirmation. The OCTO-SSR team set out to validate their findings. We experimented with custom and publicly available, simple scripting tools to test more than eighty blocklists to confirm that duplication across these was low.

[15] It is important to note that both maliciously registered and compromised domains pose security threats. A goal of this project is to represent the domain name threat landscape as it is viewed by organizational or Internet Users cannot reliably distinguish between a compromised or malicious website; instead they rely on reputation data providers to identify threats and trust to operating system or application countermeasures that employ these to protect them. .

False Positive Rates

We are sensitive to allegations or concerns regarding false positive rates; however, we believe that it is beneficial to collect and make use of the same abuse data that is reported to organizational and Internet users. We have not found academic or Internet industry studies that have found notable false-positive rates in the lists that we use, and lists we have chosen are often hand-tuned (curated) to improve accuracy[16].

Available academic literature and the extent of commercial use imply that false-positive rates are quite low among the lists we have chosen. Each list has consistently demonstrated a low false-positive rate that satisfies the risk-reward requirements of very large number of organizational subscribers or service providers (e.g., email services, ISPs) who make use of these lists. Security systems that protect billions of users incorporate these data into their threat mitigation measures If the feeds include a small number of false positives, those false positives are reflected in the DAAR system's output. However, since other parties rely on those feeds – for example, email service providers, Internet service providers, and resolver operators – the false positives will affect the domain name ecosystem regardless of how the DAAR system reports them.

Does the DAAR System Capture All of the Abuse?

The amount of abuse on the Internet associated with gTLD domain names is larger than what is cataloged by the DAAR system. The blocklist providers do not see all instances of abuse that happen on the Internet, and the providers do not list all of the domains registered by established bad actors. For these reasons, the statistics the DAAR system reports should be considered as a subset of the abuse problem in a given gTLD, or in the gTLD portfolio of a given registrar.

How Are Addition to and Removals from RBLs Handled?

Most of the abuse sources are binary: either a domain is on a list, or it is not. The abuse feeds are designed in ways that make it easy to recognize when a domain is being added to or removed from the list. Sometimes each addition to or removal from a list is an explicit event in the feed, with each event timestamped by the list operator (and DAAR stores that timestamp).

In other cases, DAAR performs a "diff" between two periodic downloads of an abuse list, and this allows the system to discover the latest additions and removals. We timestamp each add or remove for reference as we perform the diffs.

Two lists offer "confidence levels":

---

[16] See, for example, http://www.surbl.org/faqs and https://www.spamhaus.org/whitepapers/effective_filtering/

1. Every phishing URL in PhishTank's feed is tagged as either "unverified" or "verified." DAAR only uses "verified" reports, which are phishing pages that have been confirmed by two or more sources.[17]
2. The APWG feed is a streaming list of reported phishing URLs. DAAR only uses URLs that are coded with a confidence rating of 90% or higher. These are phish that have been verified either by a human, or via reliable heuristics that check the URL in question to confirm the presence of a page that is taking in data from visitors and has other tell-tale features.

The APWG phishing feed does not provide removal data. It is a stream of newly-seen phishing URLs and APWG does not attempt to track the "live" status of each URL after that. So, DAAR only counts a phish on the APWG feed as being "live" on just the day it was reported, and DAAR automatically assigns a same-day removal to the listing. This is a conservative approach – the phishing attack may be active for many days, but DAAR will only count it on the day in which it was reported.

How Does the DAAR System Attribute Spam Domains?

Blocklist providers typically extract domain names found in the bodies of spam email messages. Sometimes they also identify problematic name server names, or in the sender domain parts of email addresses, when they are able to determine with confidence that these are malicious domains and not forged. They may also include domains from URLs found in malware attachments, or URLs extracted via static and dynamic analysis or C&C and downloader malware.

The blocklist providers who examine spam use programs that analyze literally millions to billions of email messages per day. The providers decide which messages are spam, and what domains those spam messages are advertising. However, it's not possible for the blocklist providers to examine the exact content of every one of those messages, or to analyze in detail every destination domain that is being advertised. It's important to note that many of those destination domains are involved in other kinds of abuse. For example, some harbor drive-by malware, others are hosting phishing sites, and others are supporting scams of various kinds. These domains will be flagged in the "spam" category because that's how they were discovered and advertised, but they may not be flagged for other types of abuse they're being used for.

So, it's important to note that spam is not only a *category of abuse*, but it's also a vital *detection method*. The above is also a reason why the "spam" category has more domains in it than the other abuse types, and why some domains are flagged in both the spam category and another category too.

Spammers and phishers sometimes will generate thousands of subdomains on a second-level domain they buy, but subdomains are handled as described above.

[17] https://www.phishtank.com/faq.php#howmanypeoplehavetov

Is the DAAR System Attribution Mechanism Open to Gaming?

It is theoretically possible for this or any other attribution mechanism to be open to gaming. But scoring and blocking have been done for many years but we are unaware of a campaign by one party to surreptitiously damage an entire TLD's reputation. There are several reasons why the risk outweighs reward for such a campaign. One would have to invest possibly considerable funds to register a very large number of domains. The perpetrator risks the possibility of discovery or attribution during the registration of domains for the campaign. The gaming activity might need to be sustained over time to be successful. During this time, a large number the domains associated with the gaming campaign must continue to resolve and must be resilient or adaptive to having IP addresses used to send spam messages blocklisted. The perpetrator risks being discovered from the resulting damage to the hosting operators. The perpetrator risks prosecution in jurisdictions where spam has become a criminal activity. Finally, if such an operation was successful and attributed, the victim might have legal options at its disposal.

DAAR System Reporting

The objective of the DAAR system is to provide data to support ICANN community in various purposes. ICANN organizations' CEO and the Board of Directors have asked for monthly activity reporting. The ICANN organization will consult with the ICANN community for advice regarding the ways in which the reports can me most beneficially used. Several kinds of internal, stakeholder, and/or public reports or reporting are possible. Examples include:

- Periodic reports to the ICANN Board of Directors on the security threat activity for all gTLD registries or all registrars. The ICANN Board will determine whether these reports or the underlying data will be made public.
- Periodic public reports on prevalence of security threats for individual registries or registrars. In keeping with ICANN's requirements for openness and transparency, but keeping in mind potential sensitivities about the data, three approaches are under consideration:
  - Individual security threat reports will be delivered to registry or registrar operators. These may consist of aggregated statistics rather than detailed listings of specific domains[18]. The operator will have a designated period of time, e.g., 30 or 60 days, to review the report. A second report will be generated at the end of the time period and this will be made public along with reports for all registrars or registries.
  - A public report of the security threat activities for all registries or all registrars will be generated. Several characteristics or features of this report

[18] Under our current reputation data licensing agreements, ICANN can share or publish derivative data but cannot share the lists of domains.

are under consideration. These reports will show aggregated data collected on the prevalence of threats, severity of threats, and concentrations or distributions of security threats. Abuse scoring will be normalized and thus it will be more difficult for consumers of the report to deduce individual registries or registrars.

  - Lists of blocked domains. Under our current reputation data licensing agreements, ICANN can share or publish derivative data but cannot share the lists of domains. If the ICANN community wants access to lists of blocked domains from the DAAR system, ICANN must seek permission or renegotiate the reputation list providers' terms of service.

- Investigative reports. Compliance, GDD, or OCTO SSR could initiate a report when, for example, those departments become aware of extraordinary security threat activity. Such reports could be presented to an identified registry or registrar operator for compliance action or a cooperative abuse investigation with affected parties.
- Reports Solicited from the ICANN community. The ICANN community could request that DAAR data be made available for a particular, report, for example, a successor report to the Statistical Analysis of DNS Abuse in gTLDs a work product of the Consumer Confidence and Trust Review Team. In such cases, ICANN would make DAAR data available to researchers or consultants, who would conduct analyses.
- Reports Solicited by gTLD registries or accredited registrars. Registry operators or registrars could request a report to gain insight into a single event or long-term pattern of abusive behavior that is adversely affecting their operations.

Percentage of Abuse

The DAAR system currently calculates the percentage of abuse to attempt to quantify the abuse in the portfolio of any given gTLD or registrar. The DAAR system reports absolute abuse numbers for each TLD and each registrar (daily or cumulative over 365 days).

Percentage of Abuse P_{ab}: rates a gTLD or registrar based on the number of domains it currently has on the blocklists, compared to the total number of domains the registry or registrar sponsors as per the zone files. P_{ab} measures how much the registry's or registrar's portfolio is being abused at a point in time, based on what is currently considered a problem by the various data sources.

- For a gTLD, P_{ab} is the number of unique, currently listed domains per 100 domains in the zone file.

$$P_{ab} = \left(\frac{abuse\text{-}listed\ domains\ in\ TLD\ on\ this\ day}{resolving\ domains\ in\ TLD\ zone\ on\ this\ day}\right) \times 100$$

  This shows us the percentage of domains in the zone file that are currently listed on abuse blocklists that the DAAR system employs.

- For a registrar, P_{ab} is the number of unique, currently listed abuse domains per 100 domains that the registrar sponsors.

  $$P_{ab} = \left(\frac{abuse\text{-}listed\ domains\ sponsored\ by\ registrar\ on\ this\ day}{resolving\ domains\ sponsored\ by\ registrar\ on\ this\ day}\right) \times 100$$

  This shows us the percentage of the domains that the registrar sponsors that are currently listed on abuse blocklists that the DAAR system employs.

Refer to the section, DAAR Threat Data Compilation, for explanations of how we determine *abuse listed domains.*

Access to the DAAR System

The DAAR system is a custom application and subscription service developed specifically for ICANN Organization. Under our current licensing terms, only ICANN staff can access the tools. Use, report creation, and sharing of reports or data are being studied to consider the sensitivity of information, audience, and contractual constraints that ICANN must satisfy regarding the re-use of licensed data sources.

Conclusion

We are making information about the DAAR system available to the community and request input as to how to improve our approach in order to meet the goals of providing a system that has a published and vetted methodology for creating a persistent store of DNS and abuse data.

DAAR has already shown value internally within the ICANN organization in identifying unusual and unexpected DNS abuse behaviors to the OCTO SSR and Research teams. For example, the DAAR system will provide input to metrics for the Identifier Technology Health Indicators project. The DAAR system may also provide historical or snapshot data for studies conducted by parties external to ICANN organization. For example, the DAAR data allow us to observe anomalous increases or decreases in registration or abuse activity on given dates in time or over time periods. These data may be used in studies designed to attribute changes in historical behavior to external events (e.g., a spam campaign), or to promotional activity at a registrar. The DAAR data may also be used in studies to determine whether malicious actors exhibit flocking or migration behavior in their abusive domain registration activities.

To our knowledge, the DAAR system is unique in the extent of its data collection process, curation and scope. The DAAR data has the potential to support studies that have not been possible. Such studies may help the ICANN community or the Internet community gain important insights into domain name abuse that can positively influence mitigation strategies or consensus policy development.

Annex A. Identifying spam and the IP addresses and domain names that spam uses

The most commonly used and effective way to identify spam messages is to collect email messages at mail user accounts that are intentionally created to capture unsolicited email (spam). These accounts are called spam traps. Spam trap accounts are forms of honeypots. Spam trap accounts only collect email messages they receive: they do not send email or reply to email. They are not used to subscribe to or opt in to any service. Spam investigators typically construct many spam traps for collection diversity and resiliency. Email delivered to spam trap accounts can only come from spam campaigns, from email marketing campaigns that have used email addresses without consent (opt in)[19], or occasionally, from misdelivery or mistyping recipient email addresses[20].

Spam campaigns characteristically involve the transmission of millions of email messages from one or many computers to millions of email addresses. For example, Sophos spam investigators observed a week of spam activity associated with a single computer infected with malware by a spammer. The researchers discovered that:

- The infected machine sent spam to 5.5 million email addresses.
- The infected machine sent 30 gigabytes of mail.
- 26% of the email messages sent included malware (11 different malware types were identified).
- 74% of the email messages contained links to an illegal pharmaceutical website.
- The spam mails contained links that redirected recipients through 58 different hacked servers and on to an illegal pharmaceutical site.

This example not only illustrates the volume[21] of spam email that affected millions of Internet users (from a *single* infected computer) but also *the role spam plays as the delivery mechanism for security threats*. In this case the spammer hacked

---

[19] Email marketing campaigns that purchase mailing lists may be classified as spam because some or the entire purchased list has been "harvested" and the list contains one or more spam trap accounts. Most Email Service Providers (ESPs) recommend that senders not purchase mail lists in order to avoid being classified as spammers (see also MailChimp, WhatIsMyIP)

[20] Misdelivered or mistyped email is correctly identified as unsolicited but further analysis in most cases distinguishes these from spam; in particular mistypes should not appear in spam traps at the volume/numbers that spam messages would appear.

[21] Volume includes identical as well as messages that are sufficiently similar to be considered part of the same spam campaign, e.g., same message body but different Subject: line, same malware attachment, similar message body but same malicious URL, similar malicious URL but same message body, etc.

multiple servers, was infecting users with malware, and was running a botnet. This kind of malicious activity is very typical. Spammers often perform illicit activities such hacking and distributing malware, and they advertise counterfeit goods, pump-and- dump stock scams, advanced fee fraud schemes, and phishing attacks.

Spam traps identify spam senders

Properly deployed, spam traps provide the material required to identify spam messages and the IP addresses and domain names used to send them.

The headers, message bodies, and attachments of email messages collected at spam traps undergo analysis that distinguishes spam messages from legitimate messages. Spam analysis usually involves a rules-based scoring system: a set of policies or rules for composing legitimate email, and a score for each spam rule violation detected during the analysis of the message. Examples of message characteristics commonly attributed to spam are listed in Appendix 1. Among other things, these rules sniff out problematic IP addresses and domain names that were used to send the spam.

Investigators place these Universal Resource Identifiers (URIs, typically IP addresses or domain names) on blocklists. These lists are distributed as open or commercial data. Blocklists are a primary means that ISPs, email providers, companies, universities, government offices, and other entities use to protect their users from spam, malware, phishing, and other threats.

Some blocklists contain IP addresses:

- IP addresses that emit spam,
- IP network allocations (IP blocks) that are heavily populated with IP addresses that emit spam, and
- Autonomous Systems that advertise IP network allocations that are heavily populated with IP addresses that emit spam (e.g., a block possessed by a hosting provider).
- Domain names. (See below.)

Other blocklists contain domain names:

- Domains names used in sender email addresses (spam domains, often used to operate a spam email server)
- Domain names that are the destinations that the spammers are advertising. These are domains found in the links in the bodies of spam emails.
- Domain names in malicious URLs found in message bodies.
- Top Level Domains that are prevalent in sender email addresses.
  Using domain name registration data (Whois), investigators identify registrars that are prevalent sponsors of the malicious domain names they

collect during these analyses, and look for registrants who are repeat offenders.

Experienced blocklist providers have honed their techniques over the course of many years. Some examine millions of spam messages a day using advanced algorithms, plus some hands-on examination by analysts who are looking to understand the latest trends and spamming techniques. Analysts also perform research in order to attribute spam campaigns and the ownership of Internet resources to specific bad actors.

As a result, blocklists are highly dynamic. New IP addresses and domain names are constantly being added when new problems occur, and IPs and domains are constantly removed when they are no longer used for harmful activity.

Reputable blocklist providers make careful judgments about what to add and what not to add to their lists. They know that their advice will be reflected in the networks used by millions of providers and billions of users. False positives cannot be avoided entirely, but for the DAAR system, we followed the blacklist selection criteria recommended by SANS.

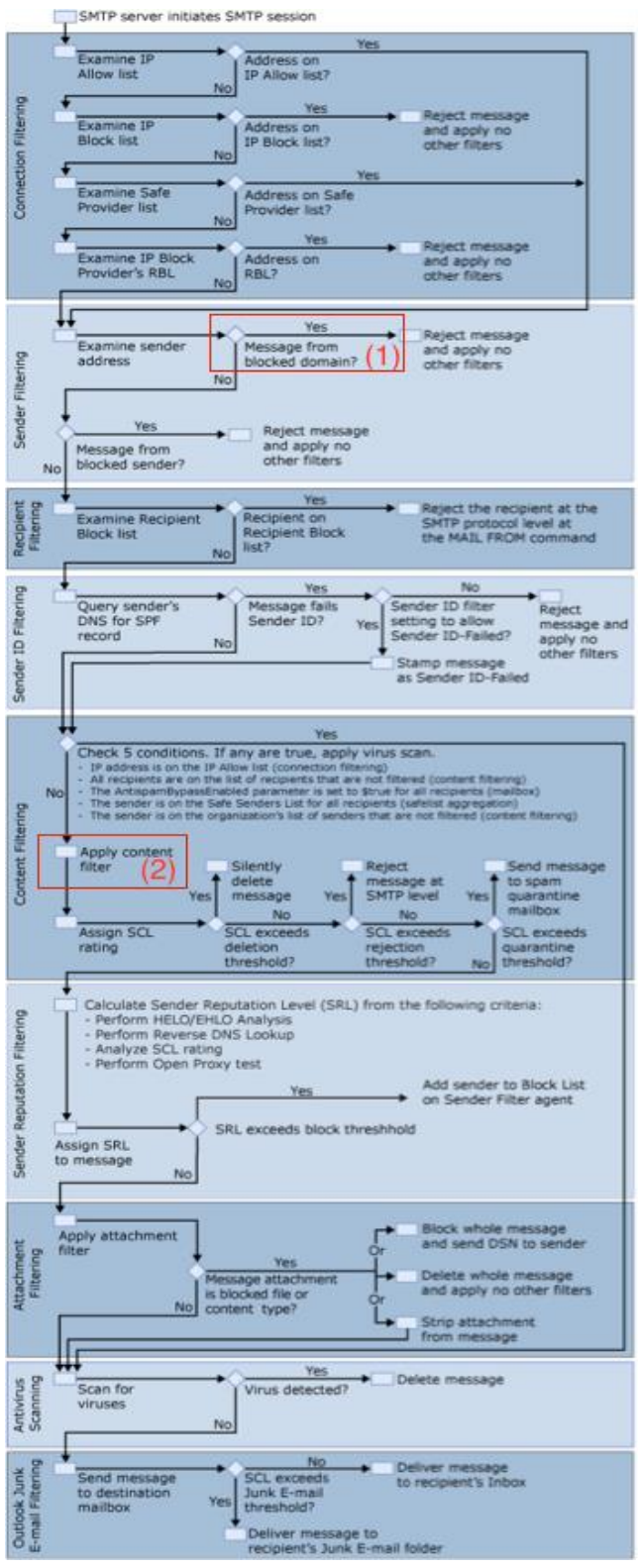
The science behind rules-based scoring is often overlooked in conversations regarding false positives. It is important, however, to put organizational and consumer confidence at the forefront of this conversation or to at least consider its value. The reason we use the same commercial or open data reputation feeds that network administrators of Email Service Providers, ISPs, government and corporate use is that these are deemed trustworthy by organizational and consumer users. They are the accepted litmus test, and thus what we report will be consistent with what organizational and consumer users see and use.

Spam email characteristics

The first column in the table below presents a partial and representative list of rules that spam emails may violate. Each of these may have a certain number of "points" associated with it, and the anti-spam system may flag an email as spam if the message exceeds a certain point threshold. Not all of these must be present in an email message to have it marked as spam, nor is any single characteristic is sufficient. The second column illustrates a representative email server spam process flow.

Spam filtering rules associated with SpamAsssassin[22]

- Connection originates on IP blocked list
- Sender address is blocked domain (1)
- Empty, missing or invalid address in To: field
- From: field same as To: field
- Received: fields trace back to a confirmed or suspected spam emitting IP address
- Confirmation that Received: fields are forged
- Comments (free form) header present
- Recipient's email address is hidden in the Bcc: field or Xreceiver field
- Message ID is missing or malformed
- Large number of recipients in To: or Cc: field
- X-Mailer: field identifies mail exchange software commonly used by spammers
- X-Distribution set to bulk
- X-UIDL: header present
- Illegal HTML or HTML comment tags present
- HTML message is present but corresponding plain text is not
- Subject: or Message body contains words or phrases common to spam
- Attachment is analyzed and found to be malware
- URL in message body is suspected or confirmed to be malicious
- Domain name in URL in message body is suspected or confirmed to be malicious (2).
- Body of email contains hidden and/or randomized text
- Links in the body of the email contain certain TLDs



[22] See http://commons.oreilly.com/wiki/index.php/SpamAssassin/SpamAssassin_Rules

Annex B. Academic Studies or Research Involving Blocklisting

Academic studies use blocklist or reputation data as the sources for studies of spam or other security threats. Some studies have evaluated the effectiveness of various blocklists, or have tried techniques to improve them. The studies and research papers listed here reinforce the fact that blocklists are a foundational means for filtering or reporting abuse.

The Statistical Analysis of DNS Abuse in gTLDs (SADAG) Report is a comprehensive DNS abuse study to analyze levels of abuse in legacy and new gTLDs, which would produce a baseline set of data for future analyses. The SADAG study serves to inform the CCTRT's analysis of potential factors explaining abuse rates in a given gTLD. The study analyzes rates of spam, phishing, and malware distribution in the global gTLD DNS from 2014 to 2016, distinguishing between legacy and new gTLDs. The methodology used in this study is similar to that used for the DAAR system

Empirically Characterizing Domain Abuse and the Revenue Impact of Blacklisting presents an economic analysis of two aspects of domain abuse in the online counterfeit drug market. Using blacklisting data, they provide an economic analysis of the revenue impact of domain blacklisting on counterfeit drugs.

Taster's Choice: A Comparative Analysis of Spam Feeds compares ten different spam feeds and finds that human identified feeds will usually be the best choice for most studies. These exhibit the best balance when managing (spamtrap) coverage, limited purity, temporal uncertainty and lack of proportionality. The authors also recommend that, "When working with multiple feeds, the priority should be to obtain a set that is as diverse as possible." We attempted to apply these findings during our selection process.

For a study, Understanding the Domain Registration Behavior of Spammers, the authors use three popularly used block lists: URIBL, SURBL, and Spamhaus DBL to explore the characteristics of registrars, domain life cycles, registration bursts, and naming patterns exhibited by spammers. The authors found that spammers employ bulk registration, that they often re-use domains previously registered by others, and that they tend to register and host their domains over a small set of registrars. The findings of their study encouraged us to consider and eventually select of Spamhaus and SURBL feeds.

In a related study, Learning to Detect Malicious URLs, the authors employed six blacklists and one white list operated by SORBS, URIBL, SURBL, and Spamhaus to develop a real time URL classification system. Our reporting system emulates or exhibits features from this study, e.g., decoupling of collection from classification and the use of large data sets from multiple reputation feeds.

As we mentioned in the main body of this paper, the findings of Metcalfe and Spring influenced our decision to employ a large number of reputation feeds, In their paper, Blacklist Ecosystem Analysis: Spanning Jan 2012 to Jun 2014, the authors demonstrated that there is little overlap among blocklists; specifically, they found that, "Based on a synthesis of multiple methods…domain-name-based indicators

are unique to one list 96.16% to 97.37% of the time…" and that, "IP-address-based indicators are unique to one list 82.46% to 95.24% of the time."

The paper, Shades of grey: On the effectiveness of reputation-based blacklists, examines the false positive and false negative rates of four block lists. The paper also explains how false positive reporting can be greatly influenced by how aggressively the spam detection thresholds are configured during experiments.

Click Trajectories: End-to-End Analysis of the Spam Value Chain is an interesting study of how the spam world works, and how to combat it.

Reputation Block Lists: Protecting Users Everywhere asserts and corroborates that reputation block lists are one of the most widely deployed and effective security solutions on the Internet. It is likely every type of entity relies on RBLs, including companies, governments, nongovernmental organizations (NGOs), mobile networks, Internet service providers, email service providers, and social networking sites.

Annex D. Operation Avalanche Infographic