DAAR Validation Report

By

John Bambenek, President

Bambenek Consulting LTD

Table of Contents

Executive Summary

This report is the final product of the analysis of Bambenek Consulting LTD on the proposed Domain Abuse Activity Reporting (DAAR) project at ICANN. Specifically, it was desired to analyze the system in the light of the following questions:

- Validate the collection and processing of zone data, domain name registration data and abuse data, or document any problems or shortcomings.
- Confirm or contest that the sources of abuse data that ICANN has selected meet industry or academic criteria for reliability, accuracy, low false positive rate, and false positive remediation. In the event that the SME refutes sources, indicate which sources and why.
- Validate the means by which the DAAR system attributes abuse domains to TLDs or document any problems or shortcomings.
- Confirm or contest that the assertions we make that spam is a security threat. We ask that the SME correct or expand assertions, if necessary. If possible, complement the existing list of academic or commercial research citations.
- Validate the generation of statistics or document any problems or shortcomings.
- Confirm that the DAAR system description is sufficiently complete for other parties to reproduce similar results using the same publicly or commercially accessible data.

By and large, the following analysis is generally supportive of this approach. By calculating what is, in effect, a “per capita” score of abuse by TLD and Registry it becomes possible to do true relative analysis on where abuse is happening more frequently and to conduct follow-on analysis on why certain TLDs or registries are being favored. Not all abuse can directly be attributed to decisions made by registries or registrars. However, there is something for “someone” to address. That someone could be a national CERT if certain regions are having more problems than others, it could be the security community at large in finding what vulnerabilities are being exploited. This data is immensely useful, and, ironically enough, follows a similar approach this analyst is using in academic research currently and independently of this project.

In reviewing a one-day window of data, it was observed some bona fide false positives were present in the data. These domains would be considered harmful if they were put into a blocklist in an enterprise environment and artificially increase a risk score of a TLD or registry. That being said, the use of “per capita” analysis helps mitigate this as it stands to reason that, by and large, the false positives will be relatively distributed across TLDs and registries in such cases. More discussion on this is below in the “Attribution to Domains” section.

However, the presence of some of these obviously benign domains (whether those domains are falsely considered to be maliciously registered or because there is something considered to be abusive on them, such as Google Drive) being scored negatively could lead some to question the validity of the approach. A truly maliciously registered domain more strongly ties the abuse to a namespace than the presence of a malicious file in a free file storage system does. Unfortunately making determinations in a programmatic way on whether a specific indicator is truly malicious, compromised, or simply a service provider being used by a criminal is well beyond the scope of DAAR. These incidents should be monitored and strategies to mitigate their inclusion should be developed.

In regard to the question on whether spam should be included as a security threat, this analyst violently agrees that it should be. The largest malware delivery mechanism is spam, spam (or at least unsolicited commercial email) is outlawed or heavily regulated in almost every jurisdiction, and much of the discussion within ICANN going on currently in regard to privacy of registrant data all point to the near-universal opinion that spam is a security and abuse threat. The specific feeds chosen to enumerate domains used in this activity largely mitigate false positives and hone in on that infrastructure that is truly abusive as opposed to merely questionable. The way these providers generate their feeds rely on delivery characteristics and not the content of messages.

This analysis largely supports the approach of data collection proposed in the whitepaper provided for this SOW. The one omission that was noted from the White Paper was how "errors" would be processed when the various feeds of abuse data were not able to be retrieved. It should be possible to handle those situations in a similar way as to when CZDS data is unavailable, namely, keep collecting and doing as much analysis as possible. In the case of feed data, it is more frequently generated so the impact of individual "errors" is more minimal. The only recommended change is to make it obvious on what days such error conditions occur so that when individuals analyze this data after the fact, it is clear from the user interface which data they should be suspect of the data returned. In conversations with the developer of DAAR, such error handling does exist, but they were not reflected in the White Paper.

Lastly, the system description is sound and makes clear not only what is attempted to be shown but an informed individual could likely reproduce a similar study without too much guesswork. The choices of the data are clear and correspond to decisions one would likely make in a cost-constrained environment. There certainly are other feeds that could be used for DNS based indicators and they can run upwards of $100,000 per year. Some other open-source feeds of domain-related data are included in "Sources of Abuse Data" below, but that data needs to be tempered with some pruning to ensure as much reduction in false positive information as possible. There is little overlap among the various open-source and low-cost intelligence feeds out there, so more data would lead to better visibility.

In conclusion, the prospect of having such a system to analyze abuse data at a per-capita level for TLDs and Registrars is exciting. While not all abuse will accrue directly to negligence or malfeasance of a registrar or registry, having such comparative data available will help lead to better understanding of the cybercriminal ecosystem and help ICANN, registries and registrars, national CERTs, and the broader security community to better tailor actions to help secure the Internet at large. I look forward to seeing this system released to the broader community and am willing to be of assistance in any way that I can to help.

Recommended Changes

The recommended changes are included below in bullet-form as they are relatively trivial to enact. The general process is sound but to increase acceptance and usability, a few adjustments should be considered.

- Create an updated document (White Paper or a separate document) that accurately reflects what is being included in DAAR. The White Paper, as written today, is outdated and the developer has made material changes.
- Removal of the Mozilla Firefox AdBlock Feed. (Note, this was mentioned in the White Paper as a source but conversations with the developer of DAAR have shown this was never included.)

- Ensure on a routine basis that “content” based feeds do not enter this list (i.e. adult websites, advertising/ad tracking domains, etc).
- Indication in the user interface when CZDS data was not completely accessible (in addition to “not plotting” data).
- Indication in the user interface when some of the source intelligence feeds were not accessible for a “long” period of time (e.g. 24 hours).
- Some adjustment for those situations when registrars have small numbers of domains under management (i.e. less than 500) where a few malicious domains can have an outsized impact to their risk score.
- Consistently reassess whether there are open-source data sources that may be valuable for this effort to increase visibility on true abuse that can be reliably attributed to domains.
- Prune Adware URLs from the feeds or remove them entirely from consideration.
- Continue to monitor for the presence of “false positive” data, what leads to its inclusion, and what, if anything, should be done to prevent it from being considered in scoring.
- Ensure that appropriate nuance is communicated to end-users of this system or products that it may produce. Not all domains in an abuse list are due to issues that can be attributed to action or non-action on a registrar/registry’s part (malvertizing, compromised domains, etc. discussed more below).
- Examine the feasibility for adding “weighting” for domains that exist over longer periods of time.
- Monitor for the rate of domains that cannot be properly attributed to a registrar due to Whois access issues (or issues with whatever successor system may be put in place).
- Continuous review for the addition of new threat feeds as relevant to his effort.

Background of Analyst

John Bambenek conducted the analysis on this project. In addition to being President of Bambenek Consulting, LTD, he works as Vice-President of Security Research and Intelligence at ThreatSTOP and is a Lecturer at the University of Illinois teaching 5 cybersecurity courses. He has 18 years' experience in cybersecurity and has spoken at conferences around the world.

What is most relevant to this analysis, is that he has produced and developed two of the largest open-source threat intelligence feeds used by thousands of organizations all over the world. The first feed is based on domain generation algorithms (DGA) that tracks over one million currently active DGA domains and can be found at http://osint.bambenekconsulting.com/feeds. The second is the Barncat Malware Configuration database that has almost 400,000 malware configurations of known-bad samples seen in the wild.

Concurrently and independent of this effort, he is leading a team of 10 graduate students using open-source information to create a globally representative analysis on TLDs, Registrars, and Hosting providers that is using a relatively similar methodology (though focusing on different outcomes and entities).

The core of his professional work revolves around the creation, curation, and operationalization of threat intelligence data to protect enterprises all over the world.

Approach

This analysis has taken several steps to analyze DAAR to provide feedback that will hopefully lead to a robust system. As part of this effort, a one-day window of a subset of the data (namely from Malware Patrol) was examined for “false positives” and other data points that might lead to results that could be questioned. In those cases, some discussion is necessary about the differences between what industry would consider false positives and what would be a false positive in this instance. For discussion on “false positives”, please see that section for how this analyst would define that term in this case. Generally, the approach of what industry would use in looking at the data is the analysis here but at key points there is more in-depth analysis of why there can and should be differences in the way ICANN and industry should look at the same data and feeds.

Some errant data was identified in this analysis, other items are questionable. For those cases, at least for the examples used for this report, querying to other third-party services was done to make a determination. By and large, the data is sound, but there are shades of grey and in some test cases, resolving the greyness to white or black was performed.

For each of the feeds referenced in the White Paper, research was done to determine how the data was gathered and the context by which it appears in the feed. In some cases, the White Paper was not specific enough to make a direct determination. In other cases, some of the data in the White Paper was not ultimately used. However, when examining the data directly, stronger conclusions have been able to be made. Generally, those feeds that focus on content-based decisions (i.e. ad tracking domains) mentioned in the White Paper were called out as unsuitable for this approach. In some cases, the Malware Patrol feeds were not able to be determined exactly how the data was collected even with direct access to the data. In those cases, “best guess” determinations were made.

In depth analysis of the various reasons a domain or hostname finding its way into blocklists was performed as that is relevant to the extent such a presence can be attributed to domain abuse and may lead some to have questions about the data and analysis.

Collection and Processing of Zone Data

As an initial matter, the use of whois to gather sponsoring registrar information may be problematic based on ongoing work with the Next-Generations RDS PDP process. As part of ongoing compliance with GDPR and other issues, access to whois data is changing. Rate-limiting ungated data, for instance, could limit the ability to use this kind of system. In some cases, entities have implied that whois will go dark entirely. That said, that system is likely changing "soon" which may require an adjustment in how registrar data is being gathered.

The initial collection of the data from CZDS or other registry sponsored process is sound as it's the primary method everyone uses and has access to. Ideally, inclusion of ccTLD zones would be a good addition to this, but that is dependent on the various ccTLDs agreeing to send the data.

In general, the method is pretty straightforward. Gather CZDS data, gather abuse data, enrich with whois data for registrar info, and perform math. This makes the DAAR system easy to understand and reproduce to verify findings as necessary.

The reporting system was pretty user-friendly and very easy to quickly derive data from this analysis. The specifics of the validity of the statistics is covered in the below section.

Error and Failure Processing

One of the omissions in the White Paper in discussing error conditions was how the unavailability (i.e. rate-limiting) of Whois might be handled. The issue was called out, but unlike the lack of TLD zone data or feed data, the absence of this data can cause some issues particularly in calculating registrar abuse scores. It is possible such unavailability could cause a skew in data such that some registrars have artificially low scores while other registrars are “fully calculated” because the Whois servers their domains may be favoring are up and available. This is somewhat mitigated by the fact that registrars for a given domain do not change “often” so only one retrieval needs to be successful.

In general, error handling for a lack of CZDS data seems to be sound. It is ambiguous in the White Paper that if a single TLD zone file is inaccessible that calculations for that day would not be done for ALL zones or just the one in question. There are pros and cons to both. If the DAAR system was only calculating TLD abuse rates, not calculating for any zone based on a single error would be extreme. Registrars, however, can operate across several TLDs so small outages can have outsized impacts. Erring on the side of not calculating data is probably the “safest” approach in this case. The White Paper doesn’t specifically call it out, but some reattempts should be made to retrieve zone data if the underlying error is merely transient in nature. Discussions with the developer of DAAR indicate there are multiple attempts to get CZDS data and validation of the data received which greatly minimizes any possibility of errors. In the case a day passes without being able to retrieve data, performing the calculations based on the last day’s data is a good workaround for this issue as the number of domains in a TLD (or under management by a registrar) would not significantly change over a given day so the abuse score should still be sound.

Error handling for when errors occur in abuse feeds is relatively sound. As feeds are generated on a more routine basis, multiple attempts can be made during a given day (and in some cases, are already done in processing), so transient errors are inherently handled. The one area of concern is that most of the feeds used for DAAR are abstracted through Malware Patrol. What this means is that individual sources in Malware Patrol could have “silent failures” yet the Malware Patrol feed itself is still working. In the short-term, such an impact is minimal, but in the long-term skewing can result. It is recommended that there is some internal tracking of indicator counts in feeds at query-time, so things can be flagged for human review (i.e. if the number of indicators for an individual source change by a significant amount).

Sources of Abuse Data

Of the five major sources included in DAAR (SURBL, Spamhaus, Phishtank, APWG, and Ransomware Tracker), all are widely used and accepted in academic and professional papers. The only concern with Phishtank is that anyone can submit URLs and some validation of that data should be performed prior to inclusion. In this case, the DAAR methodology only takes domains marked "verified" which mitigates this problem, so its inclusion is acceptable with that filter.

The remaining feeds are all provided via Malware Patrol. In some cases, there was no external reference besides Malware Patrol's website on what exactly the feed, what methodology was employed to populate it, or how they handle false positive remediation. Some of the feeds mentioned in the White Paper are either partially or entirely based on content that should not be attributed to abuse. In talking with the developer of DAAR, the only area of concern is adware / potentially unwanted application data.

For this analysis, a day worth of Malware Patrol's feed data was examined so it was possible to "work backwards" to identify potentially problematic entries and some strategies to mitigate them. Generally, these were caused by adware and that will be addressed below.

One of the Malware Patrol feeds that was selected according to the White Paper is the "Mozilla Firefox AdBlock". While there are use cases for individuals (and enterprises) to block advertising and ad-tracking related domains, they do not map to a category of abuse selected for this effort. In a similar fashion, the is another feed called "Squid Web proxy blocklist" which implies inclusion of open web proxies. Enterprises would want to block these as they are often correlated to abusive users seeking to anonymize themselves, but they might not be appropriate for this effort.

The DansGuardian feed (referenced in the White Paper) includes a variety of data that can be selected by the users and some of the data is purely content-based. For instance, one of the DansGuardian “sub-feeds” is for adult websites. Many enterprises and organizations would block such content, but its inclusion here doesn’t directly map to an abuse category absent other details (it is not unheard of for adult websites to also attempt to infect their visitors, at which point including the domain would make sense. Discussion with the developer has shown this data was not included once they developed the system.

Specifically, the White Paper calls out not using the Domain Generation Algorithm (DGA) feeds from Malware Patrol. Often when people create DGA blocklists, they include all domains that could be generated by a DGA without respect to whether a domain is actually registered. This is a valid protective approach but would include non-registered domains in DAAR statistics if included here. One DGA source may be worthwhile to include (and is listed in Appendix 1 below). This analyst produces several feeds based on DGAs available as osint.bambenekconsulting.com/feeds. Two of the feeds (dga-feed.txt and dga-feed-high.txt) are listings of all valid DGA domains but makes no attempt to see if they are registered. Those should not be considered here. There are two other feeds (c2-dommasterlist.txt and c2-dommasterlist-high.txt) which list all valid DGA domains that are currently resolving (and thus registered) within the last 3 hours (feed generated hourly) and provides some rudimentary whitelisting for known sinkholes and for domains that resolve to infrastructure that should never be blocked (the DNS root servers, 8.8.8.8, etc.). The “high” feed removes DGAs that generate short domains or use word lists for their generation as those DGAs will often “collide” with otherwise benign and independently registered domains. c2-dommasterlist-high.txt may be a good candidate for inclusion.

As a final point, there is some data bias in the feeds used for this effort that exists in a large part because some vectors of abuse do not have corresponding threat feeds. Discussed more below in detail (and to an extent in the White Paper), spam has a variety of delivery mechanisms to reach intended victims. SURBL and SpamHaus focus on email-based spam and they track it well. There are not good feeds of social media or instant messaging related indicators for spam even though they remain popular vectors to reach consumers.

As more consumers all over the world move their internet usage to mobile platforms, that will have a large impact on DNS patterns and abuse. To date, there is no real feed for mobile-based threats (malware, SMS spam, etc.). According to one recent report, the percentage of internet traffic from mobile devices is over 50% of all internet traffic[1]. That means there is a potential for a large data bias against the typical traffic patterns of Internet users. There is no recommended solution for that problem in the context of DAAR but it is an area of research interest for this analyst and potentially there will be feeds for this data in the future that may be worthy of inclusion.

[1] https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/

As more feeds become available for inclusion into DAAR, when added it can artificially skew the data compared to historical rates. For instance, mobile malware may favor a different set of TLDs than conventional malware or spam and it can lead to jarring changes in trends. This should not dissuade anyone from adding in those feeds, as the more comprehensive visibility into the threat landscape will lead to better policy outcomes. It is recommended, however, if there are additions of feed data (or subtractions of feeds, if ever relevant), there is some obvious indication of when such changes occur so that large data swings can be readily explained.

The one-day data provided does not break out which indicators belong to which sources, so detailed feed-by-feed analysis is difficult combined with the fact many don't have any public detailed description. Below is a best-effort analysis of the Malware Patrol feeds (as indicated in the White Paper) on which ones to keep, prune, or remove.

SpamAssassin: Malware URL list	Keep
Carbon Black Malicious Domains	Keep
Postfix MTA	Keep
Squid Web proxy blocklist	Remove
Symantec Email Security for SMTP	Keep
Symantec Web Security	Keep
Firekeeper	Keep
DansGuardian	Filter (make sure no content-based items)
ClamAV Virus Blocklist	Keep
Mozilla Firefox AdBlock	Remove
Smoothwall	Filter (make sure no content-based items)
MailWasher	Keep

Table 1. Malware Patrol feeds with Recommended Disposition (based on what was listed in White Paper, in some cases, DAAR developer has already removed)

Lastly, there are other open-source feeds that may be useful for inclusion to help increase the visibility into the abuse landscape. Those are included at the end of this report in Appendix I.

In addition to the open-source and free threat feeds in the appendix, there may be commercial feeds that are beneficial to consider for this project. Kasperksy and Crowdstrike, for instance, have threat feeds based on their own research into malware. To what extent they would provide reasonable access at an affordable price for this use case is not something this analysis has researched. That being said, there may be value in examining what other feeds are out there and what, if any, can be acquired at low-cost / no-cost as long as the specific data isn’t being shared back out with others. There are other intelligence sharing groups that share data for free that could lead to some interesting data sets being available but would take additional post-processing or data normalization to make accessible. There certainly are other commercial feed offerings that offer high-quality unique data, but its integration here would be subject to financial constraints to which this analyst cannot speak to.

For instance, MISP ships several open-source data sets as part of the default build and it is relatively easy to join some MISP sharing communities (NATO, CIRCL, etc.) which would yield additional hostnames or domain names that might be useful for inclusion. But in those cases, it would require setting up additional hardware and that might be beyond the scope of what is going to be done in this case.

An important note, is that many feeds will inherently have unintentional geographic biases for a wide variety of reasons. It is difficult for reasons unnecessary to go into here to create a globally representative collection of sensors to receive and process data. Attack patterns will follow geographic patterns and the biases in the feeds can lead to blind spots. Opportunistically looking for diverse geographic feeds should be attempted, but this is basically an unsolved problem for intelligence providers as well. It is being mentioned here as a potential gap of abuse data but unfortunately no specific solution can be offered except to monitor for the addition of new specifically non-Western feeds becoming available.

Complications of Adware-based Inclusion

From the one-day of “sanitized” Malware Patrol URLs provided for this analysis, there was about 146,000 different URLs in the feed. Of those, about 48,000 were flagged as AdWare. The difficulty with Adware is that there are many shades of grey from the merely noxious flavors of Adware to malicious versions. By including this category without additional filtering, some clearly wrong data may be included. It was observed, for instance, that mapquest.com was in the Malware Patrol data. Here is the specific entry:

```
MBL#25613   20060326211802 http%3A%2F%2Fcdn.mapquest.com%2Fmqtoolbarv2
9cec62b72dc705de952e28a021bff1f074a422f0ebb71683d8c3f057c428c9671830dae9
AdWare.Win32.MegaSearch.m        1668    AOL-ATDN - AOL Transit Data Ne
cdn.mapquest.com      exe
```

Virustotal detected this file with 38 of 64 engines mostly under the Adware category (or potentially unwanted programs/applications). Adware is generally annoying and many anti-virus tools will detect it as it ultimately engages in software behavior that is hard to distinguish from malicious behavior (pop-ups, injecting interstitials, monitoring user behavior). Some Adware programs probably could probably be properly labeled as malware but not all.

Windows.net was also seen with a few entries off of blob.core.windows.net with various Adware detections. Inclusion of such domains counting as abuse data could be used to attempt to discredit DAAR. Either post-filtering in a targeted way to remove such entries from consideration or removing the Adware category entirely would be appropriate to handle this situation. It the recommendation to remove this category in its entirety to be conservative and to help bolster the credibility of DAAR.

Attribution of Abuse to Domains

The key question in the DAAR project and this analysis is to what extent presence in a threat feed can be attributed to a domain, and by extension to a TLD and registrar. This will also likely be the largest area of pushback from various stakeholders.

From the perspective of how enterprises use blocklists, the following discussion matters less. Enterprises will block providers, TLDs, compromised websites and/or shared hosting providers if the threats they face from those places outweighs the harm they will endure from having a block in place. In their case, it is entirely situational, as the cost will be borne by the entity putting the block in place.

In the present case, ICANN is creating a system to analyze relative abuse statistics for domains between gTLDs and registrars and that change in context has some impact on appropriate analysis of the resulting data. Various aspects of this are described below but the main difference is that enterprises have a wide variety of actions they can take based on blocklists.

For instance, the various spam feeds usually have some form of a disclaimer saying the suitability of their feeds is tailored towards implementing the blocklists to protect mail servers and potentially web proxies (to catch users clicking on spam). Applying them to the border firewall may be an overbroad application. In the case of DAAR, the only option is to either apply a score based on an indicator or not. There is some nuance in terms of what this means for the present analysis and that is discussed below, specifically for false positives, compromised infrastructure, service providers, malvertizing, geographic considerations and statistical considerations.

Some blocklists (not included in the present study) are frequently used by enterprises for blocking malicious or suspect traffic. For instance, organizations will block IP addresses that host open UDP services like DNS or LDAP that can be used for distributed denial of service attacks. The spam feeds will often block IP addresses observed hitting their sinkholes as those IP addresses are compromised and thus can be used to send spam emails. In the present study, only hostnames and domains are considered so the risks of the above are minimized but it is important to note that it is a source of intelligence often used by blocklists and may be raised by those wanting more information about DAAR.

Additionally, some of the feeds that are ingested into DAAR are coming in URL form. Enterprises can use that data in tailored ways to minimize collateral damage. For instance, URLs could be placed into a web proxy so that a specific URL is blocked but not the domain and all other resources on it. That option isn’t available here, so some nuance in terms of what the data is saying is helpful to discuss in detail below.

Nonetheless, there may be some registrars who actively cater to criminal audiences, and DAAR can help identify such providers. However, additional analysis *must be done* before making such an assertion which is outside the scope of what DAAR can directly provide. The value of DAAR in the mind of this analyst is helping form important questions that need further research to get to the answer and there is no other tool currently available to help fill this very specific gap of domain abuse trend information.

False Positives

From a data perspective, the term “false positive” seems unambiguous. In practice for enterprises and this effort, there is a great deal of nuance in what people mean by false positives and how they are treated by those who generate feeds and those who consume them for enterprise protection. For the purposes of this analysis, the definition of false positive of this analyst is suggested to be “domains indicated as abuse domains when no such abuse is actually occurring on the domain” but it should be noted this is not the same definition industry would use.

In general, when enterprises and security companies talk about false positives, what they really mean is “false alerts” (i.e. some system indicating a security problem when there is none), which is the combination of two events: “incorrect data” in a feed, and an observation in their own network of that indicator. For instance, if an IP address for a shared business-to-consumer hosting provider is included in a threat feed (i.e. wordpress.com), most of the time that would not be considered a false positive because the likelihood of legitimate use to such a shared hosting site by an enterprise user is small and may never actually be seen.

Analyzing URLs and comparing them to malware detections seems like a sound approach to determine maliciousness of a given URL, but abstracting a URL into a hostname, and then into a domain and making a determination that the domain should be scored as abuse could be seen as problematic by some. In some cases, data gets included that probably out not to be scored as abuse (see the discussion of mapquest.com above). The specific cases of service providers, compromised infrastructure, and malvertizing are discussed at length below.

In the present case, false positives (or better yet, incorrect data) will be counted in statistics and if frequent enough, could skew some TLDs or Registrars as appearing to be used more for abuse than they really are. This is discussed below under statistical considerations. Additionally, based on the geographic biases of registrars and some gTLDs, it stands to reason that differences in the user-base may lead to relative differences between the statistics. For instance, it stands to reason that some parts of the world have more security-savvy users, more tools to help enable security, national CERTs that do better jobs at victim notification, etc., that could lead to skewing of the data based on geographic distribution of the constituent users. Some of those considerations are discussed in geographic considerations below. An important benefit of this system is that such geographic differences can be found (if they exist) to better understand the abuse ecosystem.

Ultimately one thing that is hard to discern from this approach is to what extent a domain being in an abuse feed should be attributed to the domain, and by extension, the abuse then be attributed to the TLD and Registrar. It should be noted, that enterprises are doing this attribution anyway because they do not care the underlying reasons for the abuse, they just want to block it. However, producers of malicious URL feeds would likely empathetically suggest that using very narrow data (a specific URL) to suggest abuse in a given broad namespace (a specific TLD) would not be an appropriate use of their own data.

That being said, the question of whether a domain should be listed is a subjective decision based on data that is not readily available. Some domains are registered for exclusive criminal use, others the use of DNS is incidental to the attack. That is the bulk of where "debate" could result in this implementation of DAAR. That said, DAAR is designed to determine certain things and should lead to specific follow-on research to determine why data is showing what it is.

This debate is mitigated because individual indicators do not appear to be available to end-users and that the aggregate patterns is nonetheless useful data to spur further research. If, for instance, 2 million new .com malicious domains were registered, the registrar/TLD (or reseller) may be completely blameless for that event. Nonetheless, that event and pattern is useful data to drive someone to ask, "why did the criminal ecosystem make that decision?" and answering that question with follow-on research is immensely valuable for good policy decisions both at ICANN and the broader Internet and governmental communities.

One specific case of false positives merits separate discussion and that is the registration of domains as sinkholes. A sinkhole is an otherwise malicious domain that will receive data from victims but is under active control of a "benevolent party" (government, security vendor/researcher, non-profit). These domains will often remain in an abuse feed in almost every case (except for the DGA feeds from Bambenek Consulting) because even though the domain is now "safe", victim machines beaconing to that domain still indicates compromise… just a compromise that is potentially unusable by the attacker. There are registrars that cater to sinkhole operators who will have artificially high abuse scores using this method. Notably, the Registrar of Last Resort which exists almost for the exclusive use of sinkholing. In some cases, sinkholed domains were once “bad”; in other cases, a sinkholed domain was never registered by a criminal entity. Because of the nature of the work of sinkholing, specific estimates of the size of this are difficult to obtain.

There is no good way to ferret out which domains are sinkholes versus malicious or when an otherwise malicious domain is “seized” or “transferred” to a benevolent operator. For this reason, these domains persist in the feeds for long periods of time. Further complicating this is that some sinkhole operates take significant measures to not appear to be a sinkhole. The net result is that some registrars otherwise engaging in “good behavior” can appear artificially to be worse than they are simply because they are doing “good work” in mitigating abuse by allowing sinkhole operators to use their services. This use case should specifically be prevented from degradation by the use of this data. As long as registries used the ICANN ERSR, this problem should be effectively mitigated.

Unfortunately, there is no good way to completely mitigate this data from inclusion. It is possible to use sinkdb.abuse.ch service to identify IP addresses belonging to sinkholes, use nameserver artifacts to identify them, or ask the registrars themselves (in certain cases) to identify them. But there exists no known comprehensive way to identify a domain as a sinkhole as such a technique could be also used by a criminal to mitigate sinkholing operations. This is being mentioned so appropriate nuance can be added when making conclusions solely by statistical data. In an ideal world, sinkholed domains should not be included because they are not being used in abuse. Companies may still block sinkhole traffic for a variety of reasons, but if the intent is to measure abuse in given namespaces, attempting to reduce inclusion of domains that are not being used in abuse (and cannot be again, at least in the short-term) is a worthwhile effort.

Compromised Infrastructure

Domains registered for exclusive criminal use being included in the DAAR statistical reporting is likely uncontroversial. There are a wide variety of ways enterprises deal with that situation and one current approach is to simply block all "new domains" (i.e. newly registered domains) and "newly observed domains" (i.e. when the domains first start resolving in DNS) under the relatively sound theory that most people do not immediately operationalize a domain and that the longer a domain stays "live", the more likely it is safe. Domains immediately being used (either after being registered or after first resolving in DNS) are likely because criminals want to engage in crime and derive no benefit for sitting on unused infrastructure (though certainly there are groups who register domains and do not use them for some time later to avoid the blocking mentioned above).

As a result, criminals will often use compromised domains (i.e. those domains who are primary used for legitimate uses but that have been compromised by a criminal who then use them in some method to attack others) in the chain of attacks. These sites have otherwise "good reputation" as they are, in fact, legitimate domains. Often users will deploy software that is vulnerable to attack, so criminals can simply upload files and use a compromised website for malware delivery and otherwise leave the functionality of the victim website unharmed. Some websites have vulnerable "mailer" scripts that can allow abusers to use victim infrastructure to send spam. In some cases, criminals could compromise the credentials of whatever manages

DNS for a domain and create A records otherwise unused by the victim but can be leveraged by the attacker. There are a wide variety of scenarios by which otherwise legitimate sites can be used by attackers, but they ultimately come down to two high-level cases: weaknesses in registrar-provided tools (those under the administrative control of the registrar) or weaknesses in consumer-installed tools (those installed and managed by the end user). Both situations will be discussed here.

Some registrars have fairly “bare-bones” offering and focus mostly on the conveyance of a domain to a registrant. Others will offer webhosting, email hosting, and a wide-variety of value add services. Anything that is Internet-accessible can theoretically be exploited. In general, it may be a more secure choice for the “unsophisticated” user to use registrar-provided services as that, in effect, outsources security to a larger and hopefully more security-savvy organization.

The reality is that there could be vulnerabilities in registrar-provided tools (or a vulnerability that allows bulk access to make DNS changes for the zones that have the registrar as the authorized resolver). A vulnerability in those tools could lead to sudden changes in the quantity of abuse domains attributed to a given registrar (and potentially a TLD). In this case, the abuse would not necessarily by “the fault” of the registrar, but there is a registrar-specific cause that should be discovered and remediated. The statistical report in this case is valuable to help ferret this out if it is not otherwise detected by the registrar without this tool.

In the second case, a user-deployed tool or weakness leads to widespread compromise that is then used by criminals for abusive purposes. For instance, there is recent reporting on

“Drupalgeddon 2.0” on a vulnerability that potentially could affect over one million websites[2], though presumably a smaller number of domains. In this case, there could be large scale skewing of the data depending on criminals actually exploiting these websites in a way that gets them listed in abuse feeds. The skewing could vary based on the userbase of the vulnerable software or the targeting preferences of the attacker. Registrars and TLDs would rightly be correct to take issue with such abuse listings being attributed to them. That does not, however, mean that reporting on the abuse data is useless.

If there are geographic biases in compromised infrastructure, these statistics can aid national CERTs to take specific action for their constituencies. It could help proactively identify mass-compromise events and some attributes of them (are attackers only targeting specific TLDs or specific registrars). Statistical and geographic considerations are discussed below, but the data is nonetheless valuable even for compromised domains. If patterns can be discerned for how attackers target compromised infrastructure, certain actions by ICANN, the registrar community, or others could be developed to help mitigate that. Direct attribution of the frequency of abuse domains as the fault of a registrar or registry in these cases is “unfair”, but it nonetheless represents important trend data as long as sufficient nuance is used to analyze the data.

Service Providers

In this case, “service providers” means non-registrar/TLD providers that provide some service that might be useful to a criminal and which might lead to that service ending up in an abuse feed. There are a wide variety of classes of service providers in this regard: hosting providers,

[2] https://www.securityweek.com/drupalgeddon-critical-flaw-exposes-million-drupal-websites-attacks

cloud file storage systems, file sharing systems, messaging platforms, dynamic DNS providers, DDoS mitigation or content delivery networks, etc. The specific case of malvertizing is discussed separately below.

For example, there is a variety of malware stored in Google Drive because it's availability to pseudonymously share files indiscriminately with the Internet. Attackers can either send malware to victims directly or it needs to be hosted "somewhere" and some technique employed to get the user of victim machine to download it. In many cases, combinations of both are involved. Compromised websites are one useful tool to have malware distributed to victims. Google Drive and Dropbox are also both used from time to time.

Many abuse feeds, specifically those that use URLs, will often list specific URLs to malware stored on these services. The presence, however, of google.com (for drive.google.com) or dropbox.com in abuse feeds would be considered by enterprises to be a false positive in almost every case. When considering mobile malware, the largest domains for distribution of the malware are "app stores" themselves. Including them in blocklists in an enterprise would be unwise. Feeds generated based on where apps talk to would be valuable, but no such feed exists.

It stands to reason that service providers may be relatively evenly distributed among TLDs (on a per-capita basis) or registrars with a strong favor for .com. The mitigating factor in this case is that even though abuse feeds may have many entries for Google Drive, only one domain will ultimately be counted and, it least in terms of per-capita impact in .com, the score impact is minimal. The biggest risk is for small TLDs or registrars, but it is not likely this will be much of an issue, but this cannot be said definitively. Once the system is in place, it may be worth of examining for this specific problem to see to what extent it artificially skews abuse scores

Malvertizing

In a similar way to service providers, a subset of criminals will use otherwise legitimate advertising networks to distribute malicious content to attempt to get more victims. Complicating matters is that ad tracking networks will often use similar techniques as malware operators with respect to DNS. For instance, there are two, large "user groups" of DGAs, malware operators and ad tracking networks. Both use DGAs for similar reasons. Inclusion of ad tracking networks, however, does not map to an abuse category being considered in DAAR.

Since advertising networks provide the ability of third-parties to communicate content to others, criminals will use this to communicate scams, fraud, or actual exploits to end users. The end result is that artifacts of otherwise legitimate advertising networks make their way into abuse feeds from time to time. For instance, the MapQuest data item above appears to have made it into an abuse feed due to being linked in an advertisement (which then downloaded the toolbar and saw it to be Adware which got it listed). Removing the Adware category from the Malware Patrol feed will likely help mitigate this problem.

That being said, malicious advertising *targets* (as opposed to the legitimate delivery networks themselves) should be counted as abuse; either because it is a compromised domain delivering malware or it's a domain registered for exclusive criminal use.

Geographic Considerations

Jurisdictional boundaries play a role in attack patterns, and by extension, the use of DNS by attackers. For instance, Russian cybercriminals will often not target their own country for fear of prosecution. Criminals have started reducing the use of ransomware against victims in the developed world in favor of various forms of cryptocurrency mining malware. Some tools are popular in only certain geographic areas. A wide variety of factors will lead to geographic skewing of data that will be partially reflected in gTLD and registrar abuse rates. This is especially true when considered ccTLD abuse rates, but since that is not available to DAAR at this point, it is not being considered here.

Additionally, some gTLDs are geographic in nature (i.e. gTLDs referencing specific city names) and some registrars have specific geographic targeted of their customer base. It stands to reason that the relative security posture of specific users in specific geographies widely vary and this will ultimately be reflected in the abuse scores relative to one another. It stands to reason, many people will simply look at relative abuse ranks and make "snap" judgments, having some additional data might lead to better and more nuanced policy decisions.

From a purely research standpoint, it might be useful to determine the country of the main IP address for the domain record and map that into the data with an important caveat. The presence of an IP in a certain country is, at best, marginally connected to where the user actually lives. One can buy a virtual private server in almost every country in the world regardless of where you actually live. However, the integration of geographic information in addition to registrar/TLD information might provide some interesting insights as to attack and abuse patterns. It is also true, this may “clutter the analysis” of DAAR. This analyst would be very interested in the statistical analysis of all three data points (TLD, Registrar, country code of IP address of domain) together, but ultimately that might not be in scope of DAAR. Its mentioned here simply as something to consider as geographic issues may provide some insight as to relative differences in abuse scores between TLDs and/or registrars.

Statistical Considerations

Generally, after discussion the above considerations, the statistics DAAR generates is valuable. The inclusion of other data (specifically country) might lead to richer conclusions or at least more specific lines in inquiry. As many users will simply look at Top X abuse score, some care needs to be taken that there is some nuance in interpretation. The presence of very small registrars with large abuse scores (mentioned above) is one example of how superficial views of the data can lead to unfounded decisions.

It may be worthwhile to consider differential scores (Top X improving TLDs/Registrars or Top X decreasing TLDs/Registrars) to get an idea of relative movement over time. Consideration of weighting based on domains being in a feed and unsuspended by the registrar might both be useful and overly complicating to the simplicity of DAAR calculations.

By and large, the strength of DAAR is the simplicity and soundness of how things are. The concerns expressed above revolve around the nuance of how such data should and could be used.

Spam as a Security Threat

Spam ranks as a top concern among consumers, businesses, and governments as one of the most obvious forms of abuse that is immediately visible to consumers. An entire industry has cropped up to fight spam consider the strength, vigor, and unanimity in its undesirability. Some of the largest threat intelligence feeds are almost all exclusive based from spam as it has the largest datasets. Most e-mail that transits the Internet is spam; Cisco's Talos Lab shows that 85% of emails in March 2018 were spam.[3]

It should be noted all EU countries have adopted an EU wide anti-spam law as have dozens of other countries that regulate spam to various extents. A recent indictment of Russian national Peter Levashov in the United States in connection of his operation of the Kelihos botnet specifically had counts related to email spam as being criminally actionable (Count 6 and 7).[4]

As an interesting note, all e-mail servers have "out-of-the-box" protections against spam, as do all private e-mail services. This may not seem remarkable, but web browsers have only some built-in protections against phishing sites (and only recently has it been widely deployed as part of the browser) or browser attacks, operating systems almost universally have no malware protection. The threat of spam is so severe and widely accepted that not only are their options in the applications themselves to protect against it, the default is to actually provide that protection.

[3] https://www.talosintelligence.com/reputation_center/email_rep
[4] https://www.justice.gov/opa/press-release/file/1030976/download

There are two primary ways that malware finds its way onto victim computers or inside organizations: e-mail and web-based exploit kits. Of those, e-mail is far more frequent. The attacks against the 2016 US Presidential Election and the 2017 French Presidential Election were enabled by malicious e-mail messages.

The DAAR system is using various threat feeds to populate its model to find domain-related abuse. It should be noted that the entire concept of threat feeds and what eventually became threat intelligence was being done by anti-spam organizations for a decade before it began to be used by cybersecurity companies generally. SURBL, for instance, was created in 2004. The entire notion of threat lists was created by the need and desire to prevent spam.

Ultimately the notion of abuse is socially constructed, and the consensus almost everywhere in the world is that spam is an abuse threat in that the cost to the victim is far greater than the cost to the sender. Any activity that makes the recipient expend more resources than the sender is inherently abusive.

Technically, the SMTP protocol allows for individuals to send emails simply by using IP addresses, by convention this is almost universally blocked and requires senders (malicious and otherwise) to have domain names to send messages, have MX records in DNS, and otherwise have DNS-enabled features to allow for the delivery and transmission of spam. With the exception of compromised infrastructure, most of what remains can be considered domain-related abuse as the spammer needs their domains and DNS to work in order to operate and often will register domains for exclusive malicious use.

By focusing on domain records for this analysis, there is a great deal of mitigation against false positives. For instance, a compromised website with a PHP script could be used by a criminal to send a great deal of spam, but the data in the feed would only recognize the hostname of the compromised site. That site may sit on a shared server with a million other sites, but the way the data is collected would prevent those domains from being implicated. This would not necessarily by how organizations would protect themselves in this case, but for this analysis, this treatment of the data is sound.

It should be noted that the spam feeds chosen rely on spamtraps. Essentially these are “fake email accounts” that should never get e-mail as they are not in use by legitimate users. This avoids problems of interpretation as sometimes users may believe a message is spam because they don’t like the content or a variety of other reasons. In this case, there is no “human in the mix” and the analysis and categorization are automated based on e-mail going to places it would never organically go in a legitimate case.

The White Paper does discuss spam via mechanisms other than e-mail (social media, instant messaging, etc.) and those would be valid to include in this effort. That being said, none of the feeds part of DAAR would be likely to get the information in a substantial way as the sources for those feeds are generally not messaging or social media platforms.

In short, spam should be considered an abuse threat and its inclusion in DAAR is vigorously supported. In so far as feeds or sources of data can be identified for non-email spam, those should be integrated.

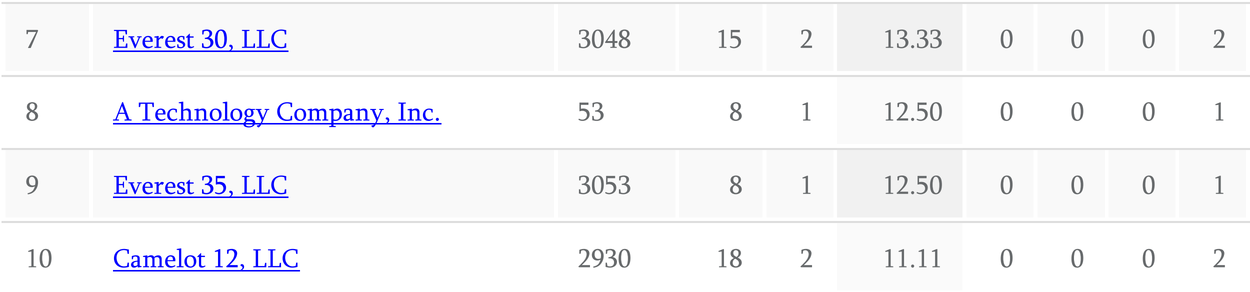
Validation of Statistics

Often, when organizations attempt to map abuse by TLDs and/or registrars, they operate in raw numbers because it is simpler to generate (especially in the case of registrars where it is not straightforward to get a full list of domains by registrar). Converting this into a percentage allows for the possibility of doing true apples-to-apples comparisons between TLDs and registrars.

The method is simple, straightforward, and hard to debate its soundness. It is exactly how this analyst can and does his own analysis on similar subjects when looking at DNS abuse data.

It may be possible to increase weighting or create another score to illustrate how long abusive domains remain operational. Purely criminal domains versus merely compromised have different lifecycles. If possible, some mechanism to analyze the time window of how long malicious domains continue to exist and comparing that across gTLDs and Registrars could lead to some useful analysis and research.

A final point on the statistics, as of this current writing, the 7th through 10th top ranked registrar for abuse each have less than 20 domains total under management.



7	Everest 30, LLC	3048	15	2	13.33	0	0	0	2
8	A Technology Company, Inc.	53	8	1	12.50	0	0	0	1
9	Everest 35, LLC	3053	8	1	12.50	0	0	0	1
10	Camelot 12, LLC	2930	18	2	11.11	0	0	0	2

Table 2. DAAR Registrar Report from April 4th, 2018 (partial listing)

What this can lead to without nuance is that registrars with small numbers of domains can be overly affected in their rankings and statistics by a single abusive domain of which they have little to no control over.

For reasons discussed in the above sections, the value in these statistics is to help hone "what questions to ask" to better understand the usage of domains in abuse. Conclusions based solely off this data will likely be less well-founded (i.e. since Everest 30, LCC is #7, it is a "bullet-proof" registrar would be ill-founded in the absence of some other data). If done properly, this data can help tailor activity to deal with the threat. Is X TLD highly-ranked because it actually intentionally provides services to criminals, some features of its business are attractive to criminals, are tools provided by the registrar vulnerable in some way, is there targeted outreach that could be done to the consumers because they are using insecure tools (i.e. CMS vulnerabilities), etc. This has the potentially of dramactically increasing the safety and security of the Internet given correct and properly nuanced follow-up based on what the statistics are saying.

System Description

The DAAR White Paper on pages 4-9 covers the description of the system and its approach to the problem. The description is thorough and complete, so another person could reproduce the results to provide verification of the outputs from DAAR. In the case of this analyst, in-depth knowledge of how this would work was already known. That being said, a moderately informed person with enough resources could reproduce a system similar to DAAR to verify the results.

Appendix 1 – Additional Feeds for Consideration

The following is a non-exhaustive list of feeds for consideration for inclusion.

https://www.alienvault.com/open-threat-exchange (requires curation)

https://hosts-file.net/?s=Download (not all feeds should be included)

https://isc.sans.edu/feeds/suspiciousdomains_Medium.txt

http://malc0de.com/database/

http://www.malwaredomainlist.com/

http://mirror1.malwaredomains.com/files/ (ie 20180404.txt)

http://osint.bambenekconsulting.com/feeds/c2-dommasterlist-high.txt

https://openphish.com/feed.txt

In addition, through the Microsoft Virus Information Alliance, it may be possible to get access to the Bing Crawler Malicious URLs feed which would provide data on over 25 million URLs daily (though all might not be worthy of inclusion.