COMMENT OF THE INTELLECTUAL PROPERTY CONSTITUENCY ON THE STATISTICAL ANALYSIS OF DNS ABUSE IN gTLDS (SADAG) REPORT

September 27, 2017

The GNSO Intellectual Property Constituency (IPC) appreciates this opportunity to comment on the data, methodology, and results of the Statistical Analysis of DNS Abuse in gTLDs (SADAG) Report. *See* https://www.icann.org/public-comments/sadag-final-2017-08-09-en. The objective of the SADAG Report was to analyze levels of abuse in legacy and new gTLDs to better understand the impact of new gTLDs on DNS abuse.

The IPC has a keen interest in the Competition, Consumer Choice, and Trust Review Team's (CCT-RT) work on the SADAG Report due to the significant overlap between online intellectual property infringement and DNS abuse, and we refer to a number of related comments by the IPC on relevant topics, including:

- IPC comments on the ICANN draft report on new gTLD program safeguards against DNS abuse;[1] and

- IPC comments on the ICANN framework of registry operators to respond to security threats.[2]

General Comments:

The IPC welcomes the SADAG report as a valuable contribution to understanding the nature of DNS abuse, and the factors which may be relevant to determining how best to address it. The IPC wishes to draw attention to several shortcomings, which the IPC acknowledges are (to a great extent) not attributable to the authors, who sought to fulfill the narrow scope of what the CCT-RT asked them to do. Nor is any criticism aimed solely at the CCT-RT, though we think their request should have been broader (to encompass IP-related abuses). Rather, in these comments we have sought to sketch out a more comprehensive program ICANN needs to undertake as part of a serious effort against abuse. That would not only include looking at a broader range of abuses (and drawing the connections between them), but also would look

[1] https://ipc.memberclicks.net/assets/ipc-position-papers/2016/2016_04april_26%20ipc%20comment%20on%20safeguards%20against%20dns%20abuse.pdf.

[2] http://www.ipconstituency.org/assets/ipc-position-papers/2017/2017_08August_2%20IPC%20Comment%20on%20Draft%20Framework%20for%20Registry%20Response%20to%20Security%20Threats.pdf.

beyond the incidence of abuse (the only focus of this study) into the effectiveness and timeliness of the responses to abuse, both by contracted parties and by ICANN compliance, and that would encompass the ccTLDs as well as the gTLDs.

To comprehensively study DNS abuse within the new gTLDs, additional research is needed to assess overlaps with trademark and copyright infringement, as well as unlawful privacy interferences and other types of abuse which implicate the illegitimate and misleading use of protected brands.[3]

The omission of IP-related abuse from the SADAG study raises serious concerns. This omission perpetuates the marginalization of IP-related abuse, instead of recognizing it as an issue that needs to be addressed and understood within the general rubric of ICANN's mission to ensure the stable and secure operation of the DNS. The absence of IP-related abuse from the SADAG Report eliminates the ability to correlate IP abuse with other types of abuse which are the focus of the study. The IPC regards the correlation between these species of abuse as significant, as evidenced by a growing body of research and the direct experience of IPC members. (The SADAG Report itself implicitly endorses this linkage by reference to the fact that phishing attacks often involve the malicious registration of strings containing trademarked terms.[4])

The IPC recognizes that additional data is essential for further research. In this regard, we encourage ICANN to support and facilitate further studies to better understand malicious web infrastructures within new and legacy gTLDs. The IPC is concerned that excluding certain forms of abuse because they have been labelled as related to "Content" is myopic because on some level, phishing and malware are themselves content-driven, and there is a growing body of evidence demonstrating a correlation between parked pages and content theft and malware distribution.[5]

The IPC also submits that a more comprehensive measurement of consumer trust and safety would look beyond comparing new gTLDs to legacy gTLDs, and also include an examination and comparison of abuse related to ccTLDs. Inclusion of abuse related to ccTLDs would provide a more complete picture from a consumer trust perspective. Miscreants simply do not respect the "type" of TLD at stake as we have seen with domain generation algorithms across many TLDs by the same miscreants.[6]

---

[3] *See*, *e.g.*, the list of threat classifications that were discussed recognized in 2012 https://www.icann.org/en/system/files/files/dns-symposium-25oct12-en.pdf.

[4] *See* SADAG Study at 12 ("For example, by manual analysis of maliciously registered domains in the fourth quarter of 2015 we find 88 abused .top domains. 75 out of 88 contain the following strings: Apple, iCloud, iPhone, their combinations, or misspelled versions of these strings suggesting that they were all used in the same phishing campaign against users of products of Apple Inc.").

[5] *See* https://www.securitee.org/files/parking-sensors_ndss2015.pdf and http://www.digitalcitizensalliance.org/news/press-releases-2016/dangerous-partners-digital-citizens-investigation-finds-that-malware-operators-and-content-theft-websites-assisted-by-u.s.based-tech-firms-are-targeting-millions-of-consumers/.

[6] For example, the DGA of the infamous Conficker botnet (version C) generated 50,000 domain names per day, which spread out over 113 TLDs.

Specific Comments:

The Report underscores the need to create a draft framework for a high security zone (HSZ) verification program through enhanced operational and security controls to establish a set of criteria to assure trust in TLDs with higher risk of being abused by miscreants. Data about Response timeframes from the time a registry receives a complaint of a malicious registration to the time of mitigation is relevant to Draft Framework for a HSZ Verification Program, and this study supports the need to renew efforts to implement this safeguard.

1. ICANN should measure the incidence and impact of registration abuse and/or malicious conducted or facilitated by contracted parties, including registrars and privacy and proxy services.

2. ICANN should prioritize investing in providing up to date reporting mechanisms to ICANN Compliance, working alongside ICANN's CTO, such as a Web-based API user interface.

3. ICANN should make available the data needed to measure not only the number of complaints, but also the resolution time, as well as by the applicable contracted parties involved, including without limitation whether false or inaccurate WHOIS is an indicator of other abusive activities that is the subject of a failure to respond to an abuse complaint or appropriately investigate it.

4. ICANN Contractual Compliance needs to specifically break down the complaints where the contracted party failed to investigate from those where it failed to respond to the abuse complaint, and keep track of how long it takes for these failures to be remedied from the time it was first reported to the contracted party, as well as from the time it was first reported to ICANN (as measured by the time of submission of the report and not from the time ICANN confirms the request is valid).

5. While ICANN may lack the ability to mandate the provision of all relevant data by contracted parties, ICANN should work with the community to obtain the data from other legitimate sources, such as the IP & brand security community, which may be willing to provide such data subject to certain agreed parameters/best practices. We encourage ICANN's CTO to work alongside ICANN Contractual Compliance to facilitate this work.

6. Section 2.15 of the Registry Agreement allows ICANN to conduct an economic study on the functioning of new gTLDs in relation to DNS abuse. Registry Operators should provide abuse reporting metrics in terms of how long it took them to confirm and mitigate threats by classification type as this bears on the economics of security in relation to DNS abuse.

7. ICANN Compliance must provide more granular reporting metrics on how long it takes them to resolve complaints that a registry or registrar failed to respond to an abuse report, and specifically how the complaint was resolved. The underlying data should be provided to support the metrics as Confidential Information (as required by Section 7.15) and ICANN should aggregate and make anonymous such data in order to provide the metrics to be

correlated with studies of DNS abuse. This should be done for legacy gTLDs to the extent the applicable agreement provides.

8. ICANN Contractual Compliance should keep metrics on how long it takes WHOIS inaccuracy report to be resolved after being reported and map the domains subject of such complaints by registrar, registry and TLD. A timeline should be put in place for implementation supervised by ICANN's CTO.

9. There should be a study of the correlation between public interest commitments voluntarily undertaken and the occurrence and persistence of abuse as well. Further, in relation to Public Interest Commitments, the IPC believes that ICANN Compliance has a positive obligation to monitor compliance with PICS in Specification 11 of the New gTLD Registry Agreement, especially as to eligibility verification about so-called "safeguard strings" where consumer protection interests are paramount. These strings include, but are not limited to ".bank, .creditunion, .mutualfunds, .attorney, .lawyer, .doctor, .charity, and numerous others as identified in ICANN NGPC Resolution No. 2014.02..05NG01 - https://www.icann.org/en/system/files/files/resolutions-new-gtld-annex-2-05feb14-en.pdf - which requires the registry to take numerous positive steps to verify that the entities registering domain names in these TLDs are qualified to do so. In the absence of regular monitoring by ICANN Compliance of these PICs, the only method of enforcement is an ad hoc complaint from a member of the community, which might be routed through the PICDRP process. Regular monitoring and random audits by ICANN Compliance will yield metrics bearing directly on the issue of maintaining Consumer Trust and Confidence in the Internet.

10. We note that Delft University of Technology, Netherlands and SIDN[7] Labs considered two types of security metrics: occurrence of abuse and persistence of abuse (based on limited data set) in Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs. Persistent of abuse seems to be missing from this study, however, and is critical for future studies. For example, the Report implies that pricing strategy (e.g., free registration and free basic hosting program) combined with a lack of a verification process of the registrant's identity may grease the wheels of domain abuse and increase immensely the number of domains registered explicitly for malicious purposes. However, it is also important to measure the length of time which the compromised phishing sites or maliciously registered domains remain active. The longer domains associated with abusive behavior remain active (days or even weeks), the greater the possibility that the gravity of the harm may not be correlated with the sheer number of abusive registrations.

11. Time to mitigation for abusive new gTLDs must be taken into account measured against both ccTLD and legacy TLD operators. We need a better understanding of persistence of abuse, which includes more transparency into abuse rate responsiveness by registrar in response to complaints, and whether there is a correlation between compliance complaints that a registrar failed to investigate or respond to an abuse complaint, and the level of abuse at the registrar.

[7] SADAG report at Footnote 33.

For example, the largest gTLD recently of the top 10 TLDs appears to be XYZ, and if they take down phishing materials quickly, the fact there is relatively a large volume becomes less relevant since the overall uptime is relatively low. There is not enough attention to time-to-remediation by new gTLD, which impacts on consumer trust and the level of exposure.

12. Thick WHOIS and centralized zone files is the foundation of the study and shows why these are essential cornerstones for brand security. In this regard, the continuing delay in implementing the Board's unanimous decision, taken in 2014, to move the two largest legacy gTLDs to a Thick WHOIS architecture is an ongoing contradiction of ICANN's stated commitment to operational excellence and to an effective multi-stakeholder process.

13. The 2013 provisions such as Section 3.18: Registrar's Abuse Contact and Duty to Investigate Reports of Abuse should be accounted for and studied statistically.

14. We caution that data obtained from abuse feeds is not an accurate reflection on the volume of abuse. The Report also fails to take into account the geographic coverage of each of the feeds, and also excludes the entire mobile landscape, such as unauthorized direct downloads and mobile spoof, which may distribute malware in order to compromise credentials of their targets. The report does not distinguish between registrars and resellers, which play an equally if not more important a role than the privacy or proxy operators.

15. ICANN should make available the data sets of grouped registries and affiliates that were relied upon in the study.

Additional Observations:

In addition to the comments above, the IPC submits the following observations regarding the SADAG Report:

- The security of a registry operator should also have been judged by the fitness of their abuse reporting mechanisms, which is not analyzed.
- The IPC notes that Spamhaus provides domain rather than URL blacklist, which means that the great majority of listed domains are maliciously registered. Spam domains also should have been defined as it is not clear if we are talking about payloads. The definition of maliciousness should be consistent across threat categories.
- The IPC believes that registrars should have been required to provide the statistics on the number of confirmed abuse complaints for malicious registrations within the dataset. If ICANN provided the number of complaints broken down further that could have been an important piece of the abuse study.
- The IPC advises that the study should have taken the hosting into account for compromised domains as the hosting variables are relevant. These variables are easily measurable at scale and capture the 'attack surface' of providers along with aspects of their business model as shown by research from Delft University, "Rotten Apples or Bad Harvest? What We Are Measuring When We Are Measuring Abuse" (cited in Footnote 38 of the SADAG Report).

- Rather than looking at absolute counts of abuse alone, measuring the amount of abuse relative to the providers' structural properties adds valuable information and should have been done by registry and registrar.
- Data-driven policy could try to improve the factors identified as influential, *e.g.*, require higher security standards at registrars offering cheap or free domains especially in bulk just like should be required by providers who host more popular websites.
- The time it takes a registrar to respond is much more relevant to health than the number of abuse instances as one could be for weeks and another for seconds. They are not equivalent.
- Lists follow similar malicious resources differently. From a practical point of view, any ten lists cannot provide a comprehensive description of all malicious indicators. Every list the defender can obtain and utilize will probably continue to provide new, non-overlapping defense to the network. Though the defender must evaluate the quality of new identifiers, any new list can provide useful identifiers of malicious activity not already contained in the defender's list. This research undermines the comprehensiveness of the blacklists used which overlapped considerably on domains when little overlap exists when taking a broader range of feeds into account. *See* Everything You Wanted to Know About Blacklists But Were Afraid to Ask (http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=83438).
- It has been well documented already that the blacklist ecosystem is intimately related to the low cost of domains and infrastructure to adversaries.
- Blacklist interrelation affects the information security evaluation and baseline creation as well. Academic and industry papers often rate performance of a particular task according to its agreement with some blacklist or lists. If all lists were equal or generation methods open, this method would be acceptable. However, because each list is different and largely non-overlapping, the ability to alter results by the choice of list leaves the evaluation process open to manipulation, as an author can choose the list that offers the best agreement.
- The 2014 blacklist study analyzed 85 blacklists and concluded that they had different algorithms leading to the blacklist that needed to be considered to fully understand its significance.
- While the report notes that it has "also considered other models that contain "Registry" as a fixed effect to capture systematic differences in the policies of registries across new gTLDs such as pricing, bulk registration options, etc. Interestingly, our results indicate that none of the registry operators have statistically significant effect on the abuse counts." The SADAG Report should provide support for this statement.
- The regression results consistently show a negative correlation between the "Type" variable reflecting strict registrations and count of phishing domains. This may be due to how phishing is reported. There should have been a study of the correlation between WHOIS inaccuracy reports and abusive domains. Not all domains are typos of brands. But a domain can still be malicious even if it was registered > 3 months ago. WHOIS inaccuracy is an indicator of other abuse. The report would have benefitted from cross-referencing the WHOIS inaccuracy reports against known blacklists.

- The best way to conduct the study in the future is to start now – measuring confirmed abuse reports by hosting companies and registrars, together with URS and UDRP decisions. The "time to mitigation" is a critical benchmark. Other forms of abuse, like counterfeiting and copyright infringement should not be ignored.
- In determining the geographical location of an abused domain, the SADAG Report authors used the address of the domain's sponsoring registrar. However, this is not a particularly reliable indicator, since registrants are free to register domains with registrars anywhere in the world (and some countries have no "domestic" registrar).
- There is no analysis in the SADAG Study of whether the authors relied on domains that actually sent spam from the payload URLs used in spam, and those payloads could be to a compromised site.
- The study focuses too much on the email spam vector without considering other vectors such as social media, malvertising, etc.

Respectfully Submitted,

Intellectual Property Constituency