Report: https://www.icann.org/public-comments/sadag-final-2017-08-09-en

The Non-Commercial Stakeholders Group (NCSG) is pleased to submit this comment in relation to the data, methodology, and results of the Statistical Analysis of DNS Abuse in gTLDs (SADAG) Report. We understand that this report is under the review of the CCT Review Team (CCT-RT). We understand that the CCT-RT will use this report as an input to evaluate the effectiveness of the nine safeguards techniques to mitigating DNS abuse following the introduction of the new gTLDs. Accordingly, this study intends to produce a comparative analysis of DNS abuse in new and legacy gTLDs while serving as a baseline for future studies.

SADAG delivers scientific analysis of quantitative data regarding DNS registrations. Properly done, such data could aid the CCT-RT in its work evaluating the effectiveness of the nine safeguards. Significant datasets were used by the authors over a defined time span.
The report offers good insights into one of the new technical security measures of the DNS, DNSSEC, by counting how many domain names were already utilising it. Other security metrics analysed include the occurrence of unique abused domains, the number of unique fully qualified domain names (FQDNs), and unique blacklisted URLs aggregated by TLDs. Furthermore, the authors make a clear distinction between compromised and maliciously registered domains so to deal with the two forms of abuse: registration and use.

NCSG Analysis of Report Conclusions

The authors drew several conclusions from their analysis, however; some of them require special attention and further investigation. Most troubling, the authors have not differentiated between the growth in DNS abuse that occurred simply because of the growth of the Internet and the number of domain name registrations, and abuse that can be specifically attributed to new TLDs or the ineffectiveness of the safeguards. This leads the authors to put an invidious burden on name space expansion — implying that expansion of the name space should somehow be expected to reduce overall levels of abuse. Any analysis of problems associated with expansion should also take into account the benefits of lower costs, wider access, and more choices.

In fact, the authors have not been able to justify some of the findings that have proved very useful to the CCT-RT in their evaluation of the effectiveness of the nine safeguards. In addition, phishing and spamming were the two types of DNS abuse activities mainly focused by the researchers' team in their conclusion, while there are many other types which can reveal more correlation to one both gTLD subset or another (legacy and new).

- **Conclusion point 1:**
    a. Authors statement: "We found that the absolute number of phishing domains has been driven by phishing domains in legacy gTLDs (mainly .com domains)."

b. Comments: This is expected and logical, since phishing tries to fool users into trusting emails it is more likely that well-established TLDs would be used. This is an important indicator that abuse is driven primarily by factors other than new TLDs, and restricting or regulating new TLDs heavily is not likely to affect certain kinds of abuse.

- **Conclusion point 2:**
    a. Authors statement: **"**While the number of abused domains remains approximately constant in legacy gTLDs, we observe a clear upward trend in the absolute number of phishing and malware domains in new gTLDs. The phishing and malware abuse rates in legacy and new gTLDs, however, are converging with time and are very similar at the end of 2016."
    b. Comments: Again, we find that this data indicates that most DNS abuse is not driven by new TLDs but by other factors related to the growth of the Internet and that restrictions and regulations on new TLDs should not be expected to affect the general level of DNS abuse.
- **Conclusion point 3:**
    a. Authors statement: "While we found higher concentrations of compromised domains in legacy gTLDs, miscreants frequently choose to maliciously register domain names using one of the new gTLDs. The registry operators of the most abused new gTLDs compete on price."
    b. Statement: We find this conclusion phrased in a biased way. It admits that legacy TLDs have higher concentrations of compromised domains, but then claims that miscreants "frequently choose" to register in new TLDs sometimes. This indicates to us that there is no significant difference between new and old TLDs. The authors affirm that it is unclear if the price is the only factor that can justify this difference, nevertheless; no other justification was explored.

- **Conclusion point 4:**
    a. Authors statement: "We also systematically analyzed how different structural and security-related properties of new gTLD operators influence abuse counts. …. Miscreants prefer to register, for example, standard new gTLD domain names, which are generally open for public registration, rather than community new gTLDs for which registries may impose restrictions on who or which entities can register their domains."
    b. Comment: This is an unsurprising outcome. The higher the costs of registration, the less likely a TLD will be used, period. Less use means also less likelihood of use by abusers. This conclusion indicates that the "newness" of the TLD has nothing to do with its susceptibility of

abuse, rather, it is the policies that govern the domain and the cost barriers.

5. **Conclusion point 5:**
    a. Authors statement: "Our findings suggest that some new gTLDs have become a growing target for malicious actors. Competitive domain registration prices, unrestrictive registration practices, a variety of other registration options such as available payment methods, free services such as DNS or WHOIS privacy, and finally the increased availability of domain names decrease barriers to abuse and may make some new gTLDs targets for cybercriminals."
    b. Comment: This conclusion fails to take into account the benefits of lower costs and easier availability. By this logic, the entire expansion of the Internet should be stopped because, as more people come online, it is likely that the absolute quantity of cybercrime will also increase. But what about the benefits of increased use and lower costs? What is the confidence interval of this conclusion? To what extent is this true? In fact, not all legacy gTLDs (18 out the total due to absence of data) were analyzed during the study and the new gTLDs were not, in the view of the NCSG,mature at the time of that the study began, therefore; related activities may be only marginal behaviors. Finally, the three-months' time span for scans may have led to undercounts, and thus, biased this conclusion.

NCSG Suggestions

With regards to the aforementioned comments, the NCSG recommends to review the report and:

1. Produce a comprehensive descriptive statistical comparison of rates of DNS abuse in new and legacy gTLDs as they pertain to spam, phishing, and malware distribution.
2. Use regression modelling to perform inferential statistical analysis, testing the correlation between passively and actively measured properties of new gTLDs as predictors of rates of abuse.

3. Analyse proportions of abusive domains across other relevant, abusive practices players, i.e. registrars and privacy/proxy service providers.

Rationale

One is cautioned about drawing early comparisons between the new and the legacy gTLDs. Given the newness of the new gTLDs, the null hypothesis of “no difference” cannot be ruled out. The report even observes that the initial growth patterns for new gTLD abuse suggest that long term patterns and rates will parallel those for legacy gTLDs.

The report observes that differential abuse rates as between new gTLDs appear related to the ease and cost of registration. In short, the abusers simply grab the low hanging fruit. It appears that there is little in the gTLD name to influence that choice. Beyond this it is also too early to know if any new approaches to abuse prevention embedded in the new gTLD procedures are having a substantial impact on abuse rates.

The results of the regression analysis need to be treated with caution. First, the incidence of abuse is a multi-factor result. Analysis needs to consider more factors and use more granularity. Also, simple regression, especially applied to Likert scale data, fails to capture the degrees of intensity, or seriousness, as between different forms of abuse.

Finally, there is good reason to endorse the GAC suggestion that ICANN should continue and expand upon the use of statistical analysis and data to measure and share information with the community information about levels of DNS abuse.

Thank you again for the opportunity to share our comments with you.