Mitigating the Risk of DNS Namespace Collisions

A Study on Namespace Collisions in the Global Internet DNS Namespace and a Framework for Risk Mitigation

Final~~Phase One~~ Report

28 OCTOBER 2015
~~4 JUNE 2014~~

TABLE OF CONTENTS

1 Preface to Final Report

4 Discussion of Public Comments and Revisions

JAS would like to sincerely thank ICANN and all of the individuals that have participated in the ICANN Community for their patience in review and comment process. We received significant and valuable feedback from the months public draft and numerous other discussions since the publication initial release of our Phase One document in February. We have amended our report. JAS, together with ICANN and Microsoft, elected accordingly. Specifically, we have addressed the following issues:

- Discussion of IPv6-related issues (see new Section 3.1.3);
- Recognition of emergent data and experience (see new text in Section 2);
- Additional discussion concerning the implementation tradeoffs of using a 127/8 IP vs. an Internet IP ("Honeypot") for Controlled Interruption (see additions to Section 3.1.7);
- Reduction of Controlled Interruption period to hold the publication of the complete Final Report until Microsoft released a fix 90 days (see new Section 3.1.2);
- Additional description of our findings regarding probability and severity of possible impacts resulting from name collision occurrences (see new text in Section 3.1.6);
- Discussion of staggered vs. consistent introduction of Controlled Interruption (see new text in Section 3);
- Recommendation to the critical MS15-011 [1] ("JASBUG") vulnerability. As a result, the overall impact of this critical vulnerability was materially reduced. collect additional logs to support long-term measurement of the collisions phenomena (see new text in Section 3.2);

Microsoft offers its appreciation to the [Coordinated Vulnerability Disclosure] CVD community and a special thanks to the reporters of the issue which has resulted in UNC Hardening: Jeff Schmidt of JAS Global Advisors, Dr. Arnoldo Muller-Molina of simMachines, The Internet Corporation for Assigned Names and Numbers (ICANN) and Luke Jennings from MWR Labs. [2]

The following pages contain updates throughout as issues related to the Microsoft vulnerability may now be discussed. Material that did not appear in the Phase One report appears in Sections 4 and 5 and Appendices A and B of this report. None of JAS' recommendations have changed.

[1] https://technet.microsoft.com/en-us/library/security/ms15-011.aspx
[2] http://blogs.technet.com/b/srd/archive/2015/02/10/ms15-011-amp-ms15-014-hardening-group-policy.aspx

- ~~Description of content that is expected to appear in our Phase Two report (see new section 3.4); and~~
- ~~Other minor modifications, improvements, and elaborations throughout.~~

~~3~~2 Summary ~~and Preface to Phase One Report~~

Collisions in the global Domain Name System (DNS) namespace have the potential to expose serious security-related issues for users of the DNS. This report dives right into the technical discussion and is targeted at readers who have been following the issue. Those new to the issue should first read the introductory documents located at: http://www.icann.org/en/help/name-collision.

We do not find that the addition of new Top Level Domains (TLDs) fundamentally or significantly increases or changes the risks associated with DNS namespace collisions. The modalities, risks, and etiologies of the inevitable DNS namespace collisions in new TLD namespaces will resemble the collisions that already occur routinely in the other parts of the DNS. The addition of multiple new TLDs over the past decade (generic and country code) has not suggested that new failure modalities might exist; rather, the indication is that the failure modalities are similar in all parts of the DNS namespace. Our research has shown that a very few root causes are responsible for nearly all collisions, and these root causes appear in nearly every classification of TLD, albeit in varying proportions.

That said, DNS namespace collisions are a complex and pervasive occurrence that manifests throughout the global Internet DNS namespace. Collisions in all TLDs and at all levels within the global Internet DNS namespace have the ability to expose potentially serious security and availability problems and deserve serious attention. While current efforts to expand the global DNS namespace have collision-related implications, the collision problem is bigger than new TLDs and must be viewed in this context.

In summary, our recommendations describe a comprehensive approach to reducing current and future DNS namespace collisions, alerting operators of potential DNS namespace related issues, and providing emergency response capabilities in the event that critical (e.g., life safety) systems are adversely impacted.

DNS namespace collisions exist outside of, and independently from, the current efforts to expand the DNS namespace. These collisions have almost certainly existed since the emergence of a global public DNS. As early as 2003, multiple researchers have pointed to the existence of queries into undelegated space received at the

root.[3,4,5,6] Our research shows that every TLD that has been added to the root since consistent data collection has occurred (2007) has exhibited some symptoms of collision activity prior to delegation.

The issue of collisions is not specific to TLDs; rather, risk exists wherever a collision crosses an administrative control boundary in the DNS. Said differently, the most dangerous DNS namespace collisions occur when *the resulting DNS query is resolved by a different administrative party than expected by the querier*. This makes intuitive sense. Because of the hierarchical nature of the DNS, the vast majority of administrative control separations occur at the TLD and Second Level Domain (2LD) levels.

Over the course of the study, JAS found no evidence to suggest that the security and stability of the global Internet DNS itself is at risk. This finding confirms the results of the *DNS Stability String Review* performed on each string during Initial Evaluation pursuant to Section 2.2.1.3.1 of the Applicant Guidebook (AGB).[7,8] The remainder of our research is focused on issues from the perspective of end-systems as consumers of the global DNS.

When faced with a range of unknowns and hypotheticals, it is important not to overlook emergent facts and experience. At the time~~As~~ we wrote the Phase One report~~write this update~~, 275 New gTLDs had~~ve~~ been delegated and over 835,000 second level registrations had~~ve~~ been added. TLDs representative of the complete range of the taxonomy JAS developed (see Section 5~~3~~.4) are represented. .berlin – a geographic term that our research suggests is heavily present in DNS search paths – has the third largest number of registrations of all new TLDs. .email and .link – short, technology-oriented generic terms that our research suggests are present in a number of hardcoded configurations – rank 6th and 7th respectively, each with over

---

[3] *Understanding DNS Evolution*, Castro, Zhang, John, Wessels, claffy, 2010, http://www.caida.org/publications/papers/2010/understanding_dns_evolution/understanding_dns_evolution.pdf

[4] *Is Your Caching Resolver Polluting the Internet?*, Wessels, 2004, http://dns.measurement-factory.com/writings/wessels-netts2004-paper.pdf

[5] *RFC 4697: Observed DNS Resolution Misbehavior*, Larson, Barber, 2006, http://tools.ietf.org/html/rfc4697

[6] *Wow, that's a lot of packets,* Wessels, Fomenkov, 2003, http://www.caida.org/publications/papers/2003/dnspackets/wessels-pam2003.pdf

[7] *gTLD Applicant Guidebook*, ICANN, 2012, http://newgtlds.icann.org/en/applicants/agb

[8] The process followed by ICANN's vendor for this review, Interisle Consulting Group, process is documented at http://newgtlds.icann.org/en/program-status/evaluation-panels/dns-stability-process-07jun13-en.pdf

30,000 2LD registrations. .company, .solutions, and .agency – terms that our research suggests are commonly hardcoded into small business-oriented configurations – are also delegated and have thousands of registrations each. Neither JAS nor ICANN is aware of even a single instance of a seriously problematic collision. Of course this fact certainly doesn't "prove the negative" but it also can't be ignored at this point.

Certainly the nature of the string impacts the drivers behind colliding behavior, and history provides lessons and data regarding the introduction of a variety of strings at the TLD. As we presented at Verisign's *Workshop and Prize on Root Causes and Mitigation of Name Collisions* (WPNC)[9] in London, strings with the potential to introduce new failure etiologies have been introduced into the TLD in the past. .pPost, (delegated in 2012) saw the most collision activity prior to delegation of any of the nine TLDs added since 2007. .pPost is interesting because "post" is also an HTTP method and a not insignificant proportion of the collisions appeared to be related to erroneous DNS lookups of text intended to be transmitted to an HTTP server. History provides lessons and data regarding the introduction of a variety of strings toat the TLD.

We believe the introduction of new TLDs offers an opportunity to educate operators regarding DNS namespace collisions and help find and remedy potential collision-related issues that may be present in their systems. As such, we recommend implementation of a 90-day "controlled interruption" period for all approved new TLDs with the exception of .corp, .home, and .mail. Registries that have not yet been delegated to the root zone shall implement controlled interruption via wildcard records; registries that have elected the "alternative path to delegation" shall implement controlled interruption by adding appropriate resource records for the labels appearing in their respective block lists. Following the 90-day controlled interruption period, registries will not be subject to further collision-related restrictions. Like the Certificate Authority (CA) revocation approach, which may be partially implemented in parallel, we believe the 90-day controlled interruption period offers a conservative buffer between potential legacy usage of a TLD and the new usage.

Lacking clear RFC 1918-like guidance directing operators to DNS namespaces safe for internal use, several such namespaces have been "appropriated" for this purpose over the years. While the etiology is subtly different, the .corp and .home TLDs are clear outliers in this respect; the use of .corp and .home for internal namespaces/networks is so overwhelming that the inertia created by such a large

[9] http://namecollisions.net

"installed base" and prevalent use is not likely reversible. We also note that RFC 6762 suggests that .corp and .home are safe for use on internal networks.[10]

Given that the Internet has demonstrated a need for RFC 1918-like DNS namespaces, we recommend that .corp and .home be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment. [11]

RECOMMENDATION 1: The TLDs .corp, .home, and .mail be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment.

Like .corp and .home, the TLD .mail also exhibits prevalent, widespread use at a level materially greater than all other applied-for TLDs. Our research found that .mail has been hardcoded into a number of installations, provided in a number of example configuration scripts/defaults, and has a large global "installed base" that is likely to have significant inertia comparable to .corp and .home. As such, we believe .mail's prevalent internal use is also likely irreversible and recommend reservation similar to .corp and .home and similarly recommend ICANN not delegate that TLD at this time.

~~RECOMMENDATION 1: The TLDs .corp, .home, and .mail be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment.~~

JAS uncovered a vulnerability not directly related to ICANN's New gTLD Program nor to new TLDs in general that has the potential to impact end-systems. Pursuant to ICANN's Coordinated Vulnerability Disclosure Process,[12] ICANN shall: "...privately disclose information relating to a discovered vulnerability to a product vendor or service provider ("affected party") and allow the affected party time to investigate the claim, and identify and test a remedy or recourse before coordinating the release of a public disclosure of the vulnerability with the reporter." Furthermore, ICANN's process states: "All parties to the disclosure generally agree to refrain from disclosing the vulnerability to the public until a remedy is identified and tested or until the threat is considered contained."

[10] *RFC 6762: Multicast DNS* (appendix G), Cheshire, Krochmal, 2013, http://tools.ietf.org/html/rfc6762

[11] RFC 6761 may be the appropriate vehicle for implementing a permanent reservation.

[12] *Coordinated Vulnerability Disclosure Reporting at ICANN*, ICANN, 2013, https://www.icann.org/en/about/staff/security/vulnerability-disclosure-05aug13-en.pdf

After extensive discussions with impacted vendors and ICANN executives, JAS is concerned that publication of the experimental methods and data contained in the complete JAS report may accelerate discovery of the vulnerability and/or serve to facilitate exploitation of the vulnerability after it is discovered. As such, pursuant to ICANN's process and out of an abundance of caution, JAS published the~~has recommended against publication of a complete~~ report in two phases: a Phase One report published in June, 2014 and~~at~~ this Final Report published after the impacted vendor addressed the vulnerability~~time~~.

~~A description of our expected Phase Two report appears in a section 3.4; the Phase Two report will be published as soon as it is prudent.~~

~~3.1~~2.1 Summary of Recommendations

RECOMMENDATION 1: The TLDs .corp, .home, and .mail be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment.

RECOMMENDATION 2: ICANN continue efforts to make technical information available in fora frequented by system operators (e.g., network operations groups, system administration-related conferences, etc.) regarding the introduction of new gTLDs and the issues surrounding DNS namespace collisions.

RECOMMENDATION 3: Emergency response options are limited to situations where there is a reasonable belief that the DNS namespace collision presents a clear and present danger to human life.

RECOMMENDATION 4: Root-level de-delegation of a production TLD is not considered as an emergency response mechanism under any circumstances.

RECOMMENDATION 5: ICANN leverage the EBERO mechanisms and functionality to respond to DNS namespace-related issues. ICANN must have the following capabilities on a 24x7x365, emergency basis: 1)~~).~~ Analyze a specific report/incident to confirm a reasonable clear and present danger to human life; 2)~~).~~ Direct the registry on an emergency basis to alter, revert, or suspend the problematic registrations as required by the specific situation; 3)~~).~~ Ensure that the registry complies in a timely manner; and 4)~~).~~ Evaluate and monitor the specific situation for additional required actions. Furthermore, we recommend that ICANN develop policies and procedures for emergency transition to an EBERO provider in the event the registry is unable and/or unwilling to comply. We recommend ICANN maintain this capability indefinitely.

RECOMMENDATION 6: ICANN require new TLD registries to publish the controlled interruption zone immediately upon delegation in the root zone. After the 90-day period, there shall be no further collision-related restrictions on the registry.

RECOMMENDATION 7: ICANN require registries that have elected the "alternative path to delegation" rather than a wildcard, instead publish appropriate A and SRV resource records for the labels in the ICANN 2LD Block List to the TLD's zone with the 127.0.53.53 address for a period of 90 days. After the 90-day period, there shall be no further collision-related restrictions on the registry.

RECOMMENDATION 8: ICANN relieve the prohibition on wildcard records during the controlled interruption period.

RECOMMENDATION 9: ICANN monitor the implementation of controlled interruption by each registry to ensure proper implementation and compliance.

RECOMMENDATION 10: ICANN work with the IETF to identify a mechanism for IPv6 that provides similar functionality to that available in IPv4's "localhost" reserved prefix.

RECOMMENDATION 11: ICANN, DNS-OARC, and the root operators explore a medium-latency, aggregated summary feed describing queries reaching the DNS root.

RECOMMENDATION 12: ICANN, DNS-OARC, and the root operators explore establishment of a single, authoritative, and publicly available archive for historical data related to the root.

RECOMMENDATION 13: ICANN explore collecting NXDOMAIN entries in DNS query logs from registry operators and contribute them to an independent data repository such as DNS-OARC for further analysis. To limit the potential for commercial gaming or use by malicious parties, we recommend that logs be provided six months in arrears.

RECOMMENDATION 14: ICANN request that the appropriate bodies further explore issues relating to collisions in existing DNS namespace, the practice of "domain drop catching," and the associated data feeds that may be leveraged by attackers when attempting to exploit collisions.

~~3.2~~2.2 Acknowledgements

JAS is grateful for the constructive engagement by numerous members of the community. We specifically want to recognize and thank:

- the Security and Stability Advisory Committee (SSAC) for thoughtful and valuable interaction while we drafted this report;
- Burt Kaliski and his team at Verisign Labs, for extensive and insightful public comments, valuable interaction with the JAS team throughout our study, and for their overall leadership on this issue including hosting the *Workshop and Prize on Root Causes and Mitigation of Name Collisions* (WPNC) in London;
- Farsight Security for contributing valuable data;
- OpenRegistry for contributing valuable data;
- Mike O'Connor for contributing valuable data; and ~~and~~
- our longtime partner simMachines for their analytical contributions.

43 Detection and Response

Since risk cannot be totally eliminated, a comprehensive approach to risk management contains some level of *a priori* risk mitigation combined with investment in detection and response capabilities. Consider fire protection; most major cities have *a priori* protection in the form of building codes, detection in the form of smoke/fire alarms, and response in the form of 9-1-1, sprinklers, and the fire department.

In terms of detecting problematic DNS namespace collisions, the initial symptoms will almost certainly appear through various IT support mechanisms, namely corporate IT departments and the support channels offered by hardware/software/service vendors and Internet Service Providers. When presented with a new and non-obvious problem, professional and non-professional IT practitioners alike frequently turn to Internet search engines for answers. This suggests that a good detection/response investment would be to "seed" support vendors/fora with information/documentation about this issue in advance and in a way that will surface via search engines when IT folks begin troubleshooting. We collectively refer to such documentation as "self-help" information. ICANN has already begun developing documentation designed to assist IT support professionals with namespace-related issues.[13]

RECOMMENDATION 2: ICANN continue efforts to make technical information available in fora frequented by system operators (e.g., network operations groups, system administration-related conferences, etc.) regarding the introduction of new gTLDs and the issues surrounding DNS namespace collisions.

One valuable suggestion from Google in the public comment period[14] is to stagger introduction of the controlled interruption~~Controlled Interruption~~ periods such that impacted parties have a reprieve between the detection and mitigation phases of their response. However, staggered controlled interruption~~Controlled Interruption~~ periods will have the side effect of causing intermittent failures, which are maddening and hard to diagnose from a system administrator perspective. Moreover, we found that systems configured in a way to create collision-related effects in the existing DNS namespaces routinely experience and tolerate intermittent failures (for example, when using a different DNS resolver) so intermittent failures are likely to resemble the status quo for impacted systems, not communicate a problem. We believe a sustained and consistent controlled

[13] *Name Collision Resources & Information*, ICANN, retrieved January 2014, http://www.icann.org/en/help/name-collision

[14] http://forum.icann.org/lists/comments-name-collision-26feb14/pdfBGWsaf8Vuk.pdf

interruption~~Controlled Interruption~~ period is the best opportunity to communicate with administrators.

However, providing advice to system administrators regarding technical mechanisms they may deploy to temporarily gain reprieve during controlled interruption~~Controlled Interruption~~ is valuable. Such advice may include the use of Response Policy Zones to temporarily rewrite query responses to something non-problematic (presumably NXDOMAIN), temporarily becoming authoritative for certain zones, etc. We recommend ICANN augment the existing technical advice to system administrators with such temporary remediation information and techniques.

It is likely that in the vast majority of expected cases, the IT professional "detectors" will also be the "responders" and any issues detected will be resolved without involving other parties.[15] However, situations in which other parties may be expected to have a role in response must be considered.

For the sake of this discussion, assume that an Internet user is experiencing a problem related to a DNS namespace collision. The term "Internet user" is intended broadly as any application, system, or device that is a consumer of the global Internet DNS. At this point in the thought experiment, disregard the severity of the problem. The affected party (or parties) will likely exercise the full range of typical IT support options available to them – vendors, professional support, IT-savvy friends and family, and Internet search. If any of these support avenues are aware of ICANN, they may choose to contact ICANN at some point. Let's further assume the affected party is unable and/or unwilling to correct the technical problem themselves and ICANN is contacted – directly or indirectly.

There is a critical fork in the road here: Is the expectation that ICANN will provide technical "self-help" information or that ICANN will go further and "do something" to technically remedy the issue for the user? We consider the options below in an escalating progression:

Option 1: ICANN provides technical support above and beyond "self-help" information to the impacted parties directly, including the provision of services/experts. Stated differently, ICANN becomes an extension of the impacted party's IT support structure and provides customized/specific troubleshooting and assistance. *We rule out this option as inappropriate and out-of-scope for ICANN.*

Option 2: At ICANN's request, referral, or direction, the registry provides technical support above and beyond "self-help" information to the impacted parties directly,

[15] Availability issues are typically detected internally whereas security issues are often detected by third parties and reported to the system operators.

including the provision of services/experts. Stated differently, the registry becomes an extension of the impacted party's IT support structure and provides customized/specific troubleshooting and assistance. *We rule out this option as inappropriate and out-of-scope for a registry.*

Option 3: ICANN forwards the issue to the registry with a specific request to remedy. In this option, assuming all attempts to provide "self-help" are unsuccessful~~not successful~~, ICANN would request that the registry make changes to their zone to technically remedy the issue. This could include temporary or permanent removal of second level names and/or other technical measures that constitute a "registry-level rollback" to a "last known good" configuration. *We consider this option feasible but undesirable as it creates considerable opportunity for operational complexities and unintended consequences. This option should only ~~to~~ be used in excessively serious circumstances.*

Option 4: ICANN initiates a "root-level rollback" procedure to revert the state of the root zone to a "last known good" configuration, thus (presumably) de-delegating the impacted TLD. In this case, ICANN would attempt -- on an emergency basis -- to revert the root zone to a state that is not causing harm to the impacted party/parties. *We consider this option feasible but even more undesirable as it creates considerable opportunity for operational complexities and unintended consequences. This option should only ~~to~~ be used in excessively serious circumstances after all previous mitigation attempts have failed.*

We note that ICANN's New gTLD Collision Occurrence Management Plan and SAC062 contemplate some of these emergency response options in a broad sense.

In any theater of operations – not just the global Internet DNS – emergency responders must be mindful of "cure is worse than the disease" scenarios wherein the response actually creates additional risks, harms, and significant potential for unintended consequences. Because of the potential operational impacts to the global Internet DNS, changes to the root zone are not to be taken lightly.

From a practical perspective, we conclude that the de-delegation of a TLD in the root would effectively be a permanent death for that TLD regardless of whether the TLD reappeared in the future.[16] This is a steep price for a registry to pay for anything but the most egregious and flagrant disregard for a serious harm.

Obviously, the severity of the harm is a critical variable. In risk analysis, severity is almost always measured economically and from multiple points of view. Any party

[16] While we note that there has always been some degree of churn in the root zone, the commercial pressures on the current new gTLDs significantly elevate the impact of a de-delegation, no matter how short.

expected to "do something" will be forced to choose between two or more economically motivated actors: users, registrants, registrars, and/or registries experiencing harm. We must also consider that just as there may be users negatively impacted by new DNS behavior, there may also be users that are dependent upon on the new DNS behavior. Unfortunately, we cannot give equal consideration to actors that are following the technical standards vs. those depending on technical happenstance or poorly implemented software for proper functionality.

Even attempting to weigh economic harm or "national security" on a global basis creates a slippery slope and forces registries and ICANN to arbitrate impossible scenarios. Concepts like "national security," "law and order," and "key economic processes" do not translate well on a global basis and risk another "Morality and Public Order" debate – which is exactly what happened when similar terms were introduced into the ICANN landscape previously. There~~Unfortunately, there~~ will not be time for such a debate in real-time, leaving emergency responders forced to make rapid decisions concerning extremely serious issues – like root-level changes – in a non-deterministic state.

Moreover, an emergency response threshold that is not well defined risks weaponization of the mechanism by commercial or government interests. Sadly, recent history has shown some governments will use a full range of tools to silence distribution of certain viewpoints over Internet channels. It is also reasonable to assume that commercial interests will attempt to "game" any mechanism for competitive advantage.

As such, we recommend that emergency response be limited to scenarios where there is a reasonable belief that the DNS namespace collision presents a clear and present danger to human life. While admittedly a high bar, we believe it is the only deterministic and non-debatable option. We feel creating a path to emergency response (including root-level changes) based on lesser factors is unwise.

Despite the previous recommendation, ICANN must prepare for the worst-case scenario. Fortunately, ICANN has already developed an emergency response mechanism as a part of the Emergency Back-End Registry Operator (EBERO) Program. The EBERO Program is designed to quickly respond to a variety of registry-level technical SLA failures; response options include an emergency (and potentially involuntary) transition of an entire registry to a new operator using a robust process that is highly scripted and exercised.

We recommend that, if necessary (in the event of an unresponsive or non-cooperative registry), a "root-level rollback" be implemented via EBERO as opposed to simply removing a TLD from the root. Shifting a registry to EBERO and making subsequent surgical changes is a superior approach to wholesale removal of an

entire production TLD – including potentially many 2LD registrations that are not causing harm.

RECOMMENDATION 3: Emergency response options are limited to situations where there is a reasonable belief that the DNS namespace collision presents a clear and present danger to human life.

RECOMMENDATION 4: Root-level de-delegation of a production TLD is not considered as an emergency response mechanism under any circumstances.

In the case of severe harm being exposed by a DNS namespace collision where the registry is unable or unwilling to take action (by altering or suspending a second level registration), ICANN could transfer the registry to an EBERO on an emergency basis and instruct the EBERO to make the required second level change to remedy the harm. While we recognize any "root-level rollback" is highly undesirable, ICANN should maintain the capability, thus ensuring that timely action can be taken in all circumstances.

RECOMMENDATION 5: ICANN leverage the EBERO mechanisms and functionality to respond to DNS namespace-related issues. ICANN must have the following capabilities on a 24x7x365, emergency basis: 1)~~).~~ Analyze a specific report/incident to confirm a reasonable clear and present danger to human life; 2)~~).~~ Direct the registry on an emergency basis to alter, revert, or suspend the problematic registrations as required by the specific situation; 3)~~).~~ Ensure that the registry complies in a timely manner; and 4)~~).~~ Evaluate and monitor the specific situation for additional required actions. Furthermore, we recommend that ICANN develop policies and procedures for emergency transition to an EBERO provider in the event the registry is unable and/or unwilling to comply. We recommend ICANN maintain this capability indefinitely.

~~4.1~~3.1 Approach to Delegation

The delegation of new TLDs presents a unique opportunity to raise awareness of the DNS namespace collision issue and help system operators identify and mitigate potential issues. Therefore, we recommend a "controlled interruption" approach as described below. The idea for controlled interruption springs from past DNS-related experiences and is conceptually similar to a "trial delegation" as proposed in SAC062.

~~4.1.1~~3.1.1 Controlled Interruption

The infamous Microsoft Hotmail domain expiration in 1999[17] and other similar domain expirations led to the implementation of ICANN's Expired Registration Recovery Policy.

More recently, Regions Bank made news[18] when their domains expired, and undoubtedly countless other similar events go unreported. In the case of Regions Bank, the Expired Registration Recovery Policy seemed to work exactly as intended – the interruption inspired immediate action and the problem was solved, resulting in only a bit of embarrassment. Importantly, there was no opportunity for malicious activity.

For the most part, the Expired Registration Recovery Policy is effective at preventing unintended expirations due to the application of "controlled interruption." The Expired Registration Recovery Policy calls for extensive notification before the expiration, then a period when "the existing DNS resolution path specified by the Registrant at Expiration ("RAE") must be interrupted" – as a last-ditch effort to inspire the registrant to take action.

Nothing inspires urgent action more effectively than service interruption.

But critically, in the case of the Expired Registration Recovery Policy, the interruption is immediately corrected if the registrant takes the required action – renewing the registration. It's nothing more than another notification mechanism – just a more aggressive round after all of the passive notifications failed. In the case of a registration in active use, the interruption will be recognized immediately, inspiring urgent action.

Like unintended expirations, DNS namespace collisions can be viewed as a notification problem. The system administrator utilizing the colliding namespace (either knowingly or unknowingly) must be notified and take action to preserve the security and stability of their systems.

Leveraging a controlled interruption to raise awareness of DNS namespace collisions draws on the effectiveness of the Expired Registration Recovery Policy with the implementation looking like a modified "Application and Service Testing and Notification (Type II)" trial delegation as proposed in SAC62. But instead of

---

[17] *Good Samaritan squashes Hotmail lapse?*, Hansen/CNET, December 27, 1999, retrieved January 2014, http://news.cnet.com/2100-1023-234907.html

[18] *Regions Bank website down, domain not renewed?*, Walsh/al.com, April 15, 2013, retrieved January 2014, http://www.al.com/business/index.ssf/2013/04/regions_bank_website_down_doma.html

responding with pointers to application layer listeners (or "honeypots"), the authoritative nameserver responds with an address inside 127/8 – the range reserved for Loopback. We recommend this approach be applied to A queries directly and MX and SRV queries via an intermediary A record (the vast majority of collision behavior observed in DITL data stems from A and MX queries).[19]

Responding with an address inside 127/8 will likely interrupt any application depending on an NXDOMAIN or some other response, but importantly also prevents traffic from leaving the requestor's host and does not facilitate a malicious actor's ability to intercede. In the same way as the Expired Registration Recovery Policy calls for "the existing DNS resolution path specified by the RAE [to] be interrupted,"", responding with a localhost reserved address should encourage immediate action by the requesting party while not exposing them to new malicious activity.

If legacy/unintended use of a DNS name is present, one could think of controlled interruption as a "buffer" or "cooling-off" period prior to use by a legitimate new registrant. This is similar to the CA Revocation Period as proposed in the New gTLD Collision Occurrence Management Plan that "buffers" the legacy use of certificates in internal namespaces from new use in the global DNS. As we discussed at ICANN Singapore, and Verisign's *Workshop and Prize on Root Causes and Mitigation of Name Collisions* (WPNC) in London, 30 to 90 day buffer periods are also commonly deployed in other large important namespaces like postal and phone numbering systems to provide feedback when changes occur. Like the CA Revocation Period approach, a set period of controlled interruption is deterministic for all parties. Unfortunately, human nature often requires a hard deadline to inspire urgent action.

Moreover, instead of using the typical 127.0.0.1 address for localhost, we recommend using a unique "flag" IPv4 address: 127.0.53.53. Because the primary objective is to communicate with system administrators through their logs, this unique and strange IP should stand out in log files, be noticed, and result in the administrator searching the Internet for assistance (we note that as of today, using Google to search for "127.0.53.53," the top 5 results are relevant). Making it known that new TLDs will behave in this fashion and publicizing the flag IP (along with self-help materials) will help administrators isolate the problem more quickly than just using the common 127.0.0.1. As hosts often have listening sockets bound to 127.0.0.1, this approach also reduces the probability of creating issues related to those servers. We also suggest that system administrators proactively search their logs for this flag IP address as a possible indicator of problems. Enterprise-wide sensors in the form of DNS query log analysis or Network Intrusion Detection Systems (NIDS) such as SNORT provide an enterprise perspective.

[19] AAAA query load suggests that collisions related to IPv6 space are far less pervasive.

Numerous experiments performed by JAS confirmed that a wide range of application layer software logs something resembling a "failed connection attempt to 127.0.53.53" which is the desired behavior. We also confirmed that all modern Microsoft, Linux, Apple, and BSD-derived operating systems correctly implement RFC 1122 (albeit with variations[20]) and keep the traffic within the host system, not transmitted over the network. This includes Linux and Windows-derived embedded operating systems. Of particular importance is Windows XP because our research has indicated that Windows XP is used extensively in industrial control and other embedded systems.

Additionally, we encourage ICANN and the IETF to work with software vendors to eventually incorporate functionality and tools to notice DNS queries that respond with this flag IP address and provide meaningful assistance. One could imagine a meaningful event in the Windows Event Log describing the situation where ~~if~~ a DNS query returns the flag IP, browsers displaying helpful diagnostic information instead of simply stating "Connection Timeout," etc.

JAS is elated that several vendors have in fact included detection and messaging around the 127.0.53.53 response. For example, recent builds of Google's Chrome browser now include the new error "ERR_ICANN_NAME_COLLISION" which provides specific and richer error messaging to the user over a general connection timeout. [21]

The ability to "schedule" the controlled interruption serves to further mitigate possible effects. One concern in dealing with collisions is the reality that a potentially harmful collision may not be identified until months or years after a TLD goes live – when a particular second level string is registered. A key advantage to applying controlled interruption to all second level strings in a given TLD in advance and at once via wildcard is that most failure modes will be identified during a scheduled time and before a registration takes place. This has many positive features, including easier troubleshooting and the ability to execute a far less intrusive rollback if a problem does occur. From a practical perspective, avoiding a complex string-by-string approach is also valuable.

The Expired Registration Recovery Policy mandates that the disruption may be for as few as eight days. However, our experiments indicate that the disruptions

[20] Some implementations route the entire /8 to localhost whereas other implementations use a host route resulting in only a /32 being dedicated to localhost. The resulting behavior during a connection attempt is slightly different, but indicative of failure in both cases.

[21] https://codereview.chromium.org/1035803003/ and https://chromium.googlesource.com/chromium/src/+/91dd3606d627036287f32bb449b09c170a0765cf

associated with controlled interruption as proposed may be more subtle, justifying a longer disruption period.

We believe the 90-day CA Revocation Period is sufficiently conservative (recall, we characterized our initial recommendation – 120 days – as "exceedingly conservative"). Given the potential seriousness of DNS namespace collisions and the immense value of detecting a harmful collision prior to a registry entering General Availability (GA), we believe the conservative approach is also warranted and recommend a 90-day controlled interruption period.

If there were to be a catastrophic impact, a surgical reversal of a 2LD registration could be implemented relatively quickly, easily, and with low risk while the impacted parties worked on a long-term solution. A new registrant and associated new dependencies would likely not be adding complexity at this point. Our recommended 90-day controlled interruption period is an ample and conservative detection and cure period for impacted parties.

Implementation of controlled interruption achieves these objectives:

- Helps notify system administrators of possible improper use of the global DNS;
- Protects these systems from malicious actors during a cure period;
- Doesn't direct potentially sensitive traffic to registries, registrars, Internet hosts/honeypots, or other third parties;
- Inspires urgent remediation action;
- Is low risk with limited opportunity for unintended consequences; and
- Is easy to implement and deterministic for all parties.

We therefore recommend controlled interruption be implemented by each new TLD registry by publishing a zone similar to the following:

```
$ORIGIN TLD
$TTL 1H
@       IN      MX 10 your-dns-needs-immediate-attention
*       IN      MX 10 your-dns-needs-immediate-attention
@       IN      SRV 10 10 0 your-dns-needs-immediate-attention
*       IN      SRV 10 10 0 your-dns-needs-immediate-attention
@       IN      TXT "Your DNS configuration needs immediate attention see URL"
*       IN      TXT "Your DNS configuration needs immediate attention see URL"
@       IN      A 127.0.53.53
*       IN      A 127.0.53.53
```

We note that some versions of popular DNS servers (notably BIND[22]) do not properly validate DNSSEC signed query responses to wildcards in all cases.

[22] *Bug 390 - NSD does not return closest provable encloser NSEC3 on wildcard queries*, NLnet Labs, May 26, 2011, retrieved January 2014, https://www.nlnetlabs.nl/bugs-script/show_bug.cgi?id=390; also note ISC RT ticket #26200

However, we also note the potential difficulties and confusion that could arise when treating the controlled interruption zones differently than production zones from an operational perspective. We have considered the tradeoffs and recommend that registries DNSSEC sign the controlled interruption zone using the same policies and procedures they intend to use when the zone is in production. A client downstream of a flawed DNS server may in some situations be "interrupted" due to the DNS server's inability to validate the signature as opposed to an interruption due directly to controlled interruption.

We recommend that the registry implement the controlled interruption period immediately upon delegation in the root zone and the prohibition on wildcard records be temporarily suspended during this period. Given the objective of controlled interruption and the reality that no registrant data will be in the zone at this point, we believe that temporarily permitting wildcard records for this purpose is not counter to established ICANN prohibitions on wildcard records and does not raise the concerns that lead ICANN to establish these prohibitions. [23]

RECOMMENDATION 6: ICANN require new TLD registries to publish the controlled interruption zone immediately upon delegation in the root zone. After the 90-day period, there shall be no further collision-related restrictions on the registry.

However, implementing a wildcard record is not prudent for a registry in GA. As such, we recommend publishing A and SRV resource records for labels in the ICANN 2LD Block List for the 90-day controlled interruption period. While arguably not an exhaustive list of queries, the 2LD Block Lists~~block lists~~ as currently constructed provide an adequate inventory[24,25] of queries sent by long-lived systems, which are the ones of most concern. The alternative – wildcard records in production zones – is less attractive and counter to established ICANN prohibitions.[26]

[23] *SSAC Report: Redirection in the com and net Domains,* ICANN Security and Stability Advisory Committee (SSAC), July 9, 2004, retrieved January 2014, http://www.icann.org/en/groups/ssac/report-redirection-com-net-09jul04-en.pdf
[24] *Public Comments on Proposal to Mitigate Name Collision Risks by Google Inc.*, Google Inc., September 17, 2013, retrieved January 2014, http://forum.icann.org/lists/comments-name-collision-05aug13/pdfkwCAlijJOp.pdf
[25] *Is Your Caching Resolver Polluting the Internet?*, Wessels, 2004, http://dns.measurement-factory.com/writings/wessels-netts2004-paper.pdf
[26] *SSAC Report: Redirection in the com and net Domains*

With the exception of .corp, .home, and .mail, this approach would apply to all registries, including the registries not eligible for the "alternative path to delegation." ICANN will make 2LD Block Lists available as required.

RECOMMENDATION 7: ICANN require registries that have elected the "alternative path to delegation," rather than a wildcard, instead publish appropriate A and SRV resource records for the labels in the ICANN 2LD Block List to the TLD's zone with the 127.0.53.53 address for a period of 90 days. After the 90-day period, there shall be no further collision-related restrictions on the registry.

RECOMMENDATION 8: ICANN relieve the prohibition on wildcard records during the controlled interruption period.

RECOMMENDATION 9: ICANN monitor the implementation of controlled interruption by each registry to ensure proper implementation and compliance.

~~RECOMMENDATION 7: ICANN require registries that have elected the "alternative path to delegation," rather than a wildcard, instead publish appropriate A and SRV resource records for the labels in the ICANN 2LD Block List to the TLD's zone with the 127.0.53.53 address for a period of 90 days. After the 90-day period, there shall be no further collision-related restrictions on the registry.~~

~~RECOMMENDATION 8: ICANN relieve the prohibition on wildcard records during the controlled interruption period.~~

~~RECOMMENDATION 9: ICANN monitor the implementation of controlled interruption by each registry to ensure proper implementation and compliance.~~

~~4.1.2~~3.1.2 Why 90 days?

By far the most prevalent public comments to our draft report were related to the 120-day controlled interruption~~Controlled Interruption~~ period. We reviewed these comments carefully and subsequently modified our thinking.

A portion of the public comment from .Club Domains, LLC sums up the issue nicely:

> *The comments of the NTAG, Donuts, Rightside/United TLD, and Ari Registry Services have thoroughly and competently explained why the 120 day interruption period of Recommendation 7 is excessively conservative. A merely conservative interruption period of 60 days is more than adequate for registries that have already been delegated, because the detrimental effects on public interest must be balanced against the security interest of a longer interruption period. A lengthened interruption period is significantly detrimental to the public interest because it would cause confusion for commercial registrants.*[27]

We like this comment because it speaks to the trade-offs between potential risks/harms and actual risks/harms. In New TLD space, controlled interruption~~Controlled Interruption~~ is a conservative mitigation against a theoretical harm. Despite a concentrated effort by a number of researchers (JAS included!) for the better part of the past two years to find actual incidences of collision-induced harms related to New TLDs, the reality is that none have been found. As of the writing of the first phase of the report~~today~~, 275 New gTLDs were~~have been~~ delegated and over 835,000 2LD registrations have been added with no indication of serious issues. As we stated earlier, while this certainly doesn't "prove the negative," the data must be taken into consideration. Based on everything we know now, the harms remain theoretical. Given no indication of actual harms, is it justifiable for JAS to recommend an "excessively conservative" and atypical duration, or is a "merely conservative" and more typical duration more appropriate? What is the tradeoff – what *actual* harms could we be causing with an "excessively conservative" approach to a theoretical harm?

[27] http://forum.icann.org/lists/comments-name-collision-26feb14/pdfEVFexxB8GK.pdf

After reviewing this issue, prior to delivery of the final Phase One report, we ~~have~~ changed our recommendation to indicate a 90-day controlled interruption~~Controlled Interruption~~ period. We have made no subsequent changes to our recommendations in in this Final report.

~~4.1.3~~3.1.3 What about IPv6?

Since IPv6 does not support a range of addresses for localhost like IPv4, there is not a straightforward analog of our controlled interruption~~Controlled Interruption~~ recommendation in v6 space. So the discussion becomes twofold: (1) is a v6 response necessary, and if so, (2) what address would be returned?

Addressing the first, we do not believe v6 responses are necessary at this time. The data we analyzed revealed a miniscule number of resolvers seeking v6-only responses (less than 1%) where the resolver doesn't appear to be dual-stacked. As of this writing, Google reports that roughly 3.5% of their users access Google over v6.[28] So while v6 adoption is certainly important and growing, v6-only hosts experiencing a DNS namespace collision do~~es~~ not appear to be a real problem today.

Regarding the second item, an address that is not a direct conceptual equivalent to 127.0.53.53 in v4 space would need to be selected (or "appropriated") for the purpose of controlled interruption.~~Controlled Interruption.~~ While experts can certainly debate this topic (we considered ::1, ::53, IP addresses within fd00::/8, fe80::/10, and ::ffff:127.0.53.53),~~) at the end of the day~~ each approach has plusses, minuses, and importantly the potential for unintended consequences. It's critical to remember that v6 implementations are comparatively young when compared to v4 implementations; the behavior of the vast majority of v4 stacks when presented 127.0.53.53 is well understood whereas the behavior of v6 implementations and their associated infrastructure when presented with ::53, fd00::53, or ::ffff:127.0.53.53 is certainly less deterministic.

So we're left with a tradeoff: do we risk potential unintended consequences of experimenting in the "fringes" of v6 for what is very likely a small benefit? Do we risk causing new problems to address what is fairly clearly a corner case? At the end of the day, we are left with no strong rationale for a v6 response and numerous reasons to be cautious of the potential for unintended consequences.

That being said, v6 support is certainly desirable in the long-term. One possible solution is working with the IETF to extend the definition of localhost to ::0/64 instead of ::1/128 to create a direct equivalent of the 127/8 space in IPv4. We recommend that ICANN work with the IETF to identify a workable long-term solution for IPv6.

[28] https://www.google.com/intl/en/ipv6/statistics.html

RECOMMENDATION 10: ICANN work with the IETF to identify a mechanism for IPv6 that provides similar functionality to that available in IPv4's "localhost" reserved prefix.

~~4.1.4~~3.1.4 Controlled Interruption Trials

In January 2015, JAS deployed the controlled interruption zone in multiple 2LD namespaces that exhibited evidence of significant collision and collision-like behavior.

As we had previously established bi-directional communication with multiple parties querying these names, we gave our contacts advance notice that we were making changes to the zone and asked them to observe and report the behavior of their systems during the controlled interruption windows.

Despite publishing phone numbers and email addresses via http and Whois, in the event the controlled interruption caused harm, not a single call or email was received. ~~Additional details of this trial will be available in a future report.~~

3.1.5 Effectiveness of Controlled Interruption

As we complete this Final Report in mid 2015, we are in a position to make limited qualitative observations about the effectiveness of controlled interruption and the impact of DNS namespace collisions in general, given the ongoing rollout of ICANN's New gTLD Program. While not a part of our initial tasking, given the timing, recording these observations here seems apt.

As of ICANN's meeting in Buenos Aires, more than 650 new gTLD strings had been delegated. These strings cover the complete range of the taxonomy JAS developed, including strings with a material volume of pre-existing collision activity such as .prod, and .app (which were not eligible for ICANN's alternative path to delegation).[29]

Over the past year, JAS has monitored technical support/discussion fora in search of posts related to controlled interruption and DNS namespace collisions. As expected, controlled interruption caused some instances of limited operational issues as collision circumstances were encountered with new gTLD delegations. While some system administrators expressed frustration at the difficulties, overall it appears

[29] http://newgtlds.icann.org/en/announcements-and-media/announcement-2-17nov13-en

that controlled interruption in many cases is having the hoped-for outcome.[30] Additionally, in private communication with a number of firms impacted by controlled interruption, JAS would characterize the overall response as "annoyed but understanding and generally positive" – some even expressed appreciation as issues unknown to them were brought to their attention.

It is worth noting that as administrators remedy the underlying DNS issues that caused them to be impacted by controlled interruption, their systems are certainly safer and likely more efficient as a result. This is a point made by several firms in contact with JAS privately.

Additionally, ICANN has received fewer than 30 reports of disruptive collisions since the first delegation in October of 2013. None of these reports have reached the threshold of presenting a danger to human life.

That being said, JAS also is aware of specific examples where controlled interruption, for whatever reason, did not cause underlying DNS issues to be remedied. Based on private contact with a party – a party JAS would consider a large and sophisticated IT operator – we learned their DNS continued to be configured in a way that caused internal DNS queries to be leaked to the Internet even following the controlled interruption period for the TLD string involved. Discussions with this party were not conclusive but JAS suspects that in this specific instance, controlled interruption was probably not disruptive enough to get the attention of operators; or if it did get the attention of operators, the issue was not viewed as important enough to cause action. Based on JAS' knowledge of the specific circumstances surrounding this operator, it is unlikely that a longer controlled interruption period or an entirely different approach to controlled interruption would have made a difference.

Controlled interruption was not expected to be perfect; few things on the Internet are. However, in general, it does appear that controlled interruption is having the expected impact and is causing at least a portion of systems with DNS namespace collision issues to be remedied. As ICANN does not require TLD operators to collect additional NXDOMAIN and controlled interruption-related logs, additional quantitative analysis and understanding of these issues will continue to be limited.

~~4.1.5~~3.1.6 Alternatives to Controlled Interruption

We considered several alternatives to controlled interruption as described above, including several honeypot approaches, use of DNAME, and various 2LD string-by-string and TLD-by-TLD approaches. While we eventually concluded that controlled

[30] http://domainincite.com/17278-victims-of-first-confirmed-new-gtld-collision-respond-fuck-google

interruption approach offers the most value and presents the least risk, discussion of alternatives is worthwhile.

~~4.1.6~~3.1.7 String-by-String Approaches (TLD and 2LD)

While the occurrence and risk associated with DNS namespace collisions is not uniform across all TLDs and 2LDs, our analysis concluded that any collision and any harm could – at least in theory – occur anywhere in the global DNS namespace. We found ample evidence supporting this conclusion, and found that it would be a quixotic undertaking to determine the root cause of every incidence of a DNS namespace collision.[31] With the exception of .corp, .home, and .mail, which are clear outliers for the reasons mentioned earlier, the several root causes we found are not limited to particular strings, new or existing TLDs, or even specific levels of the DNS.

JAS' assessment is, with the exception of .corp, .home, and .mail, that the risk of a collision in the newly applied-for TLD namespaces causing more than a highly localized disruption is low after the recommended mitigation technique is applied. String-by-string and TLD-by-TLD approaches add significant complexity and potential for unintended consequences while adding little if any security value. Not a good tradeoff. As such, we recommend an approach that address the root causes and does not delineate between specific strings unnecessarily.

~~4.1.7~~3.1.8 Honeypot Approaches

Significant discussion has occurred in several fora regarding various implementations of a trial delegation that directs traffic to an Internet-based honeypot. The honeypot, run by ICANN or some trusted third party, could serve two functions: 1) Present helpful information for operators reaching the site over http and potentially other protocols; and 2) Collect logs to help identify volume, sources, and potential severity of collision and collision-like activity. Some ideas describe a honeypot that runs for a deterministic time period while others continue the honeypot until some threshold is achieved, indicating risk has been mitigated to an (undefined) acceptable level.

Because collisions are largely a notification problem, we like the concept of honeypot approaches. However, there are some critical traits of honeypot approaches that make them undesirable.

- Whenever logs are collected, the question "for what purpose" must be asked. How much collision activity is "OK" – what is the acceptable risk? Is the threshold the same for all TLDs? Are all query sources to be treated equally –

[31] *Focused Analysis on Applied-For gTLDs - .cba*, Verisign Inc., September 15, 2013, retrieved January 2014, http://forum.icann.org/lists/comments-name-collision-05aug13/msg00039.html

that is, do we look differently upon log entries that *appear* to be from a nuclear power plant vs. a residential broadband network? These questions, being subjective in nature, may not have answers that can achieve consensus.

- Whenever logs are collected, we must also be vigilant for gaming opportunities. Because there are many interested parties and significant commercial pressures, we assume that competing interests will attempt to exploit any activity that may create an argument for slowing or halting valuable registrations in a TLD. Even the possibility (perceived or actual) of such gaming will virtually assure that gaming occurs.
- There are collision scenarios where returning an Internet IP address will cause traffic to be sent over the Internet that was never previously sent. Ever conscious of "cure being worse than the disease" concerns, we certainly do not want to open these hosts to new risks while we try to help them. Additionally, we are informed by the vulnerability we discovered on this matter; for machines impacted by the issue, honeypotting a popular port will *assure* that sensitive information is transmitted in the clear over the LAN and the Internet to the honeypot. Absent the honeypot, transmission of this sensitive information is not assured. Controlled interruption should not *decrease* the security posture of a system, even temporarily. Or, as Verisign cleverly said in their public comment, we don't want to risk turning "Controlled Interruption" into "Controlled Exfiltration!"[32]
- As security researchers have long known, a lot of potentially sensitive information appears in logs. Usernames and passwords regularly appear in http logs. Other protocols raise similar concerns. Our experience confirms that any advertised honeypot IP will receive a host of sensitive information. Managing this information -- and convincing the global Internet community that the data is being handled responsibly -- is another hurdle with any honeypot approach.
- Different global legal jurisdictions place restrictions on data collected after it is~~was~~ "solicited." As advertising a honeypot IP could be argued as "soliciting traffic," the resulting data may have legal protections, further adding to the complexity.
- Very limited experience exists related to large-scale honeypotting of the service discovery protocols and corporate directory protocols that dominate colliding DNS queries. SAC06[33] contains a lengthy discussion of the unintended consequences of these sorts of interactions with non-HTTP protocols. There is sufficient risk of causing collateral damage and unintended consequences.

[32] http://forum.icann.org/lists/comments-name-collision-26feb14/pdffjLkIlhcj4.pdf

[33] https://www.icann.org/en/system/files/files/report-redirection-com-net-09jul04-en.pdf

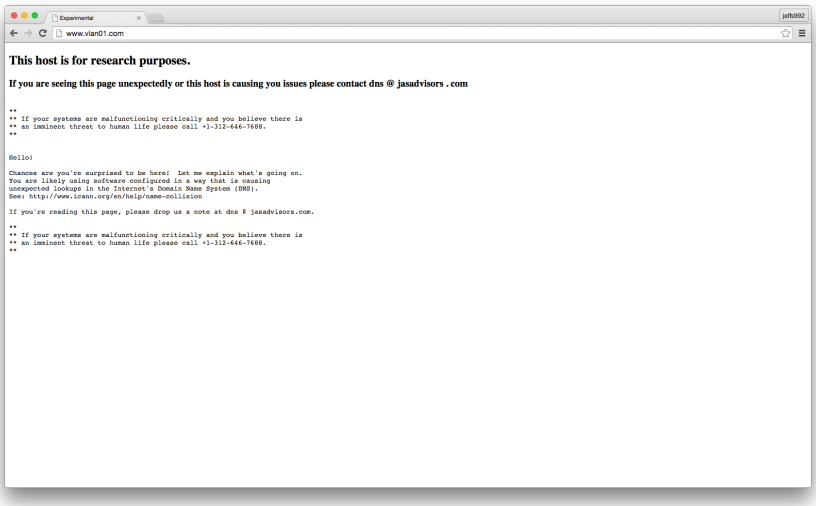
The final four bullets describe our rationale for a 127/8 IP address that does not cause traffic to leave the host, thereby avoiding those pitfalls.

We also considered a variation wherein the honeypot would be an RFC 1918 IP address as opposed to an Internet address – thereby allowing private network operators to monitor and capture the resulting traffic. However, we ruled out this variation due to the potential for unintended consequences if the RFC 1918 IP happened to be in-use in the network where the affected party resides, and because of the potential for causing general confusion. An operator with the requisite sophistication to redirect or capture RFC 1918 traffic likely also has the requisite sophistication to react appropriately to 127/8 responses.

From our research, we learned that much of the offending application layer traffic related to DNS namespace collisions is not user/"eyeball" HTTP rendered in a browser. In retrospect, this makes sense: a large proportion of the non-random-label traffic is still machine-generated related to Microsoft Active Directory, Bonjour, and Web Proxy Auto-Discovery Protocol (WPAD) protocols to name a few. Messaging via an HTTP honeypot – however useful and well-intended – is unlikely to ever be viewed by a human.

We tested this theory by hosting a number of HTTP honeypots in known high-collision second level registrations, an example of which can be found here:

http://www.vlan01.com/



Even though all of our HTTP honeypot pages contained the overt request to contact us, JAS received not a single notification. Reviewing our HTTP logs, less than 8% of DNS resolutions ultimately led to the retrieval of one of our HTTP honeypot pages.

Reviewing the HTTP logs further, less than 12% of those 8% reported an HTTP user-agent that could be considered a user-facing application (i.e. a Browser).

The other benefit of a honeypot is the data it generates for future analysis. However, as we saw with the JASBUG[34], a honeypot would also generate a potentially long-lived target list of hosts or domains with vulnerable DNS configurations, vastly increasing the risk (and potential liability) to the operator of such a honeypot.

Additionally, there is a specific scenario related to JASBUG we are concerned about: there is an insidious downgrade attack where a Microsoft Active Directory Member and Domain Controller can be tricked into communicating security-critical data including Group Policy Objects, SMB mount points, and login scripts over unencrypted and unauthenticated HTTP (WebDAV). This is information that would be communicated with an HTTP honeypot over the Internet in the clear if such a "controlled interruption" honeypot were implemented. This becomes a clear example of what Verisign correctly worried about as "Controlled Exfiltration" [35] and would cause information to be transmitted over the Internet in the clear that very likely would not be transmitted absent this honeypot. This risks causing a net reduction in the security posture of the impacted systems – the very systems we're attempting to help.

As such, JAS believes the risk of operating such a honeypot does not justify the value and recommends against such an implementation.

~~4.1.8~~3.1.9 DNAME Approaches

We considered multiple schemes using DNAME records in an attempt to emulate similar controlled interruption behavior. While we ~~eventually~~ concluded that these schemes are not feasible and are less effective than localhost-based ideas, discussion is worthwhile.

One option could be implemented via DNAME records in the root. We quickly considered this option infeasible due to the difficulties, unknowns, and potential for unintended consequences surrounding the placement of DNAME records in the root; furthermore, such an approach is very likely incompatible~~not compatible~~ with the IANA/Verisign/NTIA root zone management system as currently implemented and might~~may~~ require modifications to the IANA Functions contract.

[34] https://www.jasadvisors.com/jasbug-improper-use-of-the-dns-for-authentication-and-over-the-internet-exploitation-scenarios/
[35] http://forum.icann.org/lists/comments-name-collision-26feb14/pdffjLkIlhcj4.pdf

However, using wildcards in the delegated zone is a more viable option and emulates most of the desired behavior.

Consider a wildcard DNAME record within the origin of the TLD zone pointing to some identifiable target (e.g., "you-need-to-change-your-dns-config-see-collisions-dot-icann-dot-org."). The target should not be resolvable in order to force an NXDOMAIN response (note that this assumes the specific DNAME implementation returns an NXDOMAIN instead of SERVFAIL or something else – given the relative newness of DNAME in the DNS protocol suite and its lack of significant exercise in implementations, unusual implementation decisions and/or behavior can't be ruled out).

When considering DNAME approaches, client support is a paramount concern. While the experiments[36] conducted by Geoff Huston and George Michaelson are valuable and informative, they are biased to heavy clients and human browsing (running Flash and receiving ads). The situation before us is far less biased to these types of clients, so client support is in question at best. Proper support of DNAME (RFC 2672 circa 2000) in legacy, possibly misconfigured, devices is probably less likely than proper localhost support (RFC 1122 circa 1989).

DNAME-based approaches do offer additional flexibility when compared to localhost redirection approaches, specifically in the ability of sophisticated operators to observe, control, and redirect the responses. But again, an IT operation sophisticated enough to control DNAME queries ~~certainly~~ has plenty of other options available to manage DNS namespace collisions. Catering to sophisticated IT operators by providing flexibility and options seems to come at the expense of simplicity, predictability, and widespread client support.

Finally, DNAME-based approaches don't necessarily interrupt, negating the whole purpose of controlled interruption. The DNAME redirect to return NXDOMAIN means folks can continue on as they're currently doing. They won't notice anything so they won't fix it, defeating the purpose of the interruption.

As such, we consider DNAME-based approaches inferior to localhost-based approaches.

---

[36] *draft-jabley-dnsop-as112-dname-01: AS112 Redirection using DNAME*, Abley, Dickson, Kumari, Michaelson, October 12, 2013, retrieved January 2014, http://tools.ietf.org/html/draft-jabley-dnsop-as112-dname-01 (see *Appendix A: Assessing Support for DNAME in the Real World*)

~~4.2~~3.2 Root Level Data, Monitoring, and Day-In-The-Life (DITL)

We blogged[37] about our experiences using the DNS-OARC-maintained "DITL" datasets; these datasets are truly invaluable albeit limited for researchers looking into global Internet DNS traffic. Conscious of the calls for additional datasets and monitoring at the root level, we want to discuss the objectives of monitoring and logging systems at a meta level.

When considering monitoring and logging systems, one must always start with the "for what purpose" questions. Different data consumers have different requirements. For example, operators interested in emergency response demand a low-latency, actionable, "ticket" type of monitoring. They want the "this hard drive is dead" ticket as soon as possible after it dies. Capacity planners want intermediate-latency data with some ad-hoc aggregation and trending capabilities to answer questions like "how much data do we have and what is the growth rate?" Product managers want high-latency, highly detailed data repositories that can answer a full range of complex ad hoc queries to observe behaviors, trial new product ideas, etc.

Obviously, these very different consumers have very different requirements driving very different technical implementations.

We observe that from an availability standpoint, low-latency ticket/availability data is already available for the root. Albeit in a highly decentralized fashion, the DNS root is probably one of the most highly monitored systems on Earth in that regard.

Conversely, DITL datasets are at the other end of the spectrum: extremely high latency (one 50 hour period annually), voluminous and unstructured data suitable only for compute-intensive ad hoc analysis by expert researchers.

While individual root operators certainly have a full range of data available to them, there is nothing in the middle available to researchers or the Internet at large.

Looking from a slightly different angle, the *availability* and *content* of the root is exceptionally well monitored with low latency but the *queries* to the root are much less visible.

We believe there is a need for a medium-latency, aggregated, and more "consumable" data stream from the root operators containing aggregated summary data describing the queries seen by the root. This new feed should be in a reasonably accessible and well-documented format like CSV, XML, or YAML and

---

[37] *Demystifying DITL Data [Guest Post]*, Kevin White, JAS Global Advisors LLC, November 16, 2013, retrieved December 2013, http://domainincite.com/15068-demystifying-ditl-data-guest-post

ideally have latency on the order of a few days. Mindful of the numerous issues surrounding such an undertaking, we recommend that ICANN, DNS-OARC, and the root operators explore such a mechanism.

We note ongoing efforts by the Root Server System Advisory Committee ("RSSAC") to address monitoring, and the forthcoming publication of RSSAC 002: *Recommendations on Measurements of the Root Server System.* We applaud the proactive efforts of some root operators to increase the fidelity of root server monitoring.

RECOMMENDATION 11: ICANN, DNS-OARC, and the root operators explore a medium-latency, aggregated summary feed describing queries reaching the DNS root.

Over the course of our research, we were also surprised to find that authoritative historical information regarding the contents of the root zone is not always available. A significant proportion of historical information is only captured informally in email threads and in the heads of various luminaries. As such, we also recommend that a single, authoritative archive for root data be established.

RECOMMENDATION 12: ICANN, DNS-OARC, and the root operators explore establishment of a single, authoritative, and publicly available archive for historical data related to the root.

We recognize that data and measurement regarding the DNS namespace collision phenomenon is important. One of the attractive features of a honeypot approach is that it provides a new, high fidelity, and low latency data stream describing this behavior. In lieu of the honeypot, we recommend ICANN explore collecting NXDOMAIN and controlled interruption~~Controlled Interruption~~ (127.0.53.53 query response) entries in DNS query logs from registry operators and contribute them to an independent data repository such as DNS-OARC where they may be analyzed by the research community. To limit the potential for commercial gaming or use by malicious parties, we recommend that logs be provided six months in arrears. If such logs are collected commencing with delegation, the long-term effectiveness of controlled interruption~~Controlled Interruption~~ may be measured; we believe it important to be informed by these metrics when considering future mitigation techniques in delegated and un-delegated DNS namespace.

RECOMMENDATION 13: ICANN explore collecting NXDOMAIN and controlled interruption~~Controlled Interruption~~ (127.0.53.53 query response) entries in DNS query logs from registry operators and contribute them to an independent data repository such as DNS-OARC for further analysis. To limit the potential for commercial gaming or use by malicious parties, we recommend that logs be provided six months in arrears.

~~4.3~~4 Collisions in Existing DNS Namespace

Because of the popularity of .com, typical software behavior, and common DNS search path configurations/practices, collisions at the 2LD level within .com likely occur at a higher frequency than collisions at any other location in the DNS (2LD and TLD). Because of the sheer size and prevalence of .com, this is not unexpected. With respect to collisions, .com is a victim of its own success. Recently, other researchers have quantified the order of magnitude of collisions within .com using different datasets.[38] Noted security researcher Robert Stucke spoke at DEFCON 21 about vulnerabilities he discovered by leveraging DNS namespace collisions within .com.[39]

Researching collisions in existing TLD namespaces was a part of our engagement. Over the course of this study, JAS registered several 2LDs to enhance our understanding of this phenomenon and collect additional data. Based on behaviors uncovered during our research, we made educated guesses as to where problematic collisions may occur. These registrations immediately generated a surprising amount of traffic.

It is worth noting that while selecting 2LDs to register for our research, we made use of publica~~l~~ly available tools designed to facilitate "domain drop catching" and various "squatting" activities. One such tool offers to the public the ability to find 2LDs within .com that are "available with traffic" – the very definition of a DNS namespace collision – at the second level within the Internet's most popular TLD.

While we understand the commercial value of this service, as security practitioners we are deeply concerned about this type of functionality. As such, we recommend that ICANN request that the appropriate bodies (GNSO, SSAC, etc.) further explore issues relating to collisions in existing DNS namespace, the practice of "domain drop catching," and the associated data feeds that may be leveraged by attackers when attempting to exploit collisions.

RECOMMENDATION 14: ICANN request that the appropriate bodies further explore issues relating to collisions in existing DNS namespace, the practice of "domain drop catching," and the associated data feeds that may be leveraged by attackers when attempting to exploit collisions.

[38] http://forum.icann.org/lists/comments-name-collision-05aug13/pdf056yDnxGje.pdf

[39] http://www.youtube.com/watch?v=ZPbyDSvGasw

4.4 Malware~~Description of Forthcoming Phase Two Report~~

~~JAS uncovered a vulnerability not directly related to ICANN's New gTLD Program nor to new TLDs in general that has the potential to impact end-systems. In fact, the vulnerability manifest while researching collisions within .com while analyzing collisions in existing TLDs (please see below). Pursuant to ICANN's Coordinated Vulnerability Disclosure Process,[40] ICANN shall: "...privately disclose information relating to a discovered vulnerability to a product vendor or service provider ("affected party") and allow the affected party time to investigate the claim, and identify and test a remedy or recourse before coordinating the release of a public disclosure of the vulnerability with the reporter." Furthermore, ICANN's process states: "All parties to the disclosure generally agree to refrain from disclosing the vulnerability to the public until a remedy is identified and tested or until the threat is considered contained." As such, pursuant to ICANN's process and out of an abundance of caution, JAS has recommended against publication of a complete report at this time.~~

~~The Phase Two report is expected to contain the following information:~~

4.1 ~~Impact of malware~~/adware/click fraud~~clickfraud~~ tools

~~:~~ We found that malware, adware, and click fraud~~clickfraud~~ tools generated a significant proportion of the observed random and pseudo-random string queries~~;~~. ~~We also identified other potential sources of algorithmic~~ queries~~. In the Phase Two report, we will describe this occurrence and provide specific details and supporting data. It is worth noting that we previously discussed this finding in presentations in ICANN Buenos Aires, ICANN Singapore, and Verisign's *Workshop and Prize on Root Causes and Mitigation of Name Collisions* (WPNC) in London.[41] Queries~~ related to malware/adware/click fraud~~clickfraud~~ tools explain in excess of 20% of the total~~colliding~~ queries returning NXDOMAIN~~revealed~~ in DITL datasets.

Referring to Appendix B, a large proportion of the total number of DNS queries observed in DITL datasets contain seemingly random, pseudo-random, machine-generated or otherwise linguistically nonsensical qname components. In some TLDs, as much as 20-30% of the observed queries fall into this category. The vast majority of the random and pseudo-random labels we detected are likely related to:

- Win32/Protector malware family (multiple variants)

~~[40] *Coordinated Vulnerability Disclosure Reporting at ICANN*, ICANN, 2013, https://www.icann.org/en/about/staff/security/vulnerability-disclosure-05aug13-en.pdf~~

~~[41] http://namecollisions.net~~

- Conficker/W32.Downadup (multiple variants use random 8-11 character and random to 4-9 character DNS labels) [42]
- Kraken and Torpig botnets utilize more complex label generation algorithms as described in the paper referenced in the footnote [43]
- Multiple click fraud and search optimization toolkits

It's worth noting that the Google AdSense network now prohibits generation of random numbers within ad code to limit the potential for click fraud.[44] This policy seems to have been implemented by Google roughly in 2010 and we expect other ad networks behave similarly at this point.

While not malware/adware or a click fraud tool, the Google Chrome browser is responsible for a significant proportion of the observed DNS queries containing a random label. When Chrome browser is started, it attempts to connect to three random domains like http://jdwhvnehaz/ or http://odheucnlwq/. This of course generates the requisite DNS queries (and related queries due to local operating system DNS search path processing and DNS path devolution). The result is that a set of DNS queries containing a random 10 character label is transmitted at Chrome startup (and in some circumstances, routinely on an ongoing basis).[45] When observed from the perspective of the Internet DNS, the initial three queries generated by Chrome often result in more total DNS activity due to local operating system DNS search path processing.

The stated purpose of these queries in Chrome is to detect the presence of an upstream DNS server that may be rewriting NXDOMAIN responses. We note that other software may utilize a similar tactic and be responsible for additional randomized queries.

Noting that "random detection" is an imperfect art, the above items in aggregate account for nearly 80% of the random and pseudo-random labels we detected~~**Analysis of Collisions in previous TLD delegations:** We found that collisions have occurred prior to delegation of every TLD since (at least) 2007 and presented high-level data to this effect at the WPNC in London and ICANN Singapore. Said differently, collisions in the DNS namespace are certainly not a new phenomenon (please also see Section 2 above). In the Phase Two report, we will further describe this occurrence and provide specific details and supporting data.~~

[42] http://mtc.sri.com/Conficker/
[43] http://www.ece.tamu.edu/~reddy/papers/tnet12.pdf
[44] https://support.google.com/adwordspolicy/answer/176108
[45] https://groups.google.com/a/chromium.org/forum/ - !topic/chromium-discuss/F70-k_PGhEg

~~Analysis of Collisions in existing TLDs:~~ ~~As stated in Section 3.3 above, JAS is very concerned about collisions in existing DNS namespace and the tools that facilitate discovery of colliding names. In the Phase Two report, we will describe these phenomena, including the methods we used to discover the vulnerabilities, locate vulnerable hosts, and datasets where appropriate.~~

~~A Taxonomy of Queries and TLDs:~~ ~~We classified behavior leading to collisions and created a high-level description of each applied-for TLD based on the colliding queries present~~ in DITL datasets and in excess of 41% of the total NXDOMAIN traffic described in the DITL datasets. This is consistent with the observation that the "Alternate Path to Delegation" Second Level Domain (SLD) Collision Block Lists published by ICANN are comprised largely of these seemingly random, pseudo-random, machine-generated or otherwise linguistically nonsensical labels. [46]

[46] http://newgtlds.icann.org/en/announcements-and-media/announcement-2-17nov13-en

5 Etiology of DNS Namespace Collisions

Observing DNS queries that generate NXDOMAIN responses is only half of the story. It is important to understand what software is generating these queries and why. To gain insight into this, we needed to observe application layer network traffic related to the DNS queries.

Because DNS namespace collisions are not unique to any specific top level domain – new or legacy – we found the easiest mechanism for observing application layer traffic was to purchase "colliding" second level registrations on the open market and direct those DNS registrations and resulting application layer traffic to servers we control. Over the course of this study, JAS registered in excess of fifty (50) 2LDs using the open market and received support from others – notably the owner of corp.com Mike O'Connor – who graciously provided data and/or delegated domains to JAS temporarily in support of this research.

To select the domains for registration, we used a combination of freely available tools, results of our DITL analysis, and our own intuitions concerning underlying behaviors as the study progressed and we collected additional data. As mentioned previously, we made use of publicly available tools designed to facilitate "domain drop catching" and various "squatting" activities – particularly one such tool that offers to the public the ability to find 2LDs within .com that are "available with traffic" (a DNS namespace collision by definition). As we became more informed about collision behavior, we were able to leverage this tool to find troubling registrations like vlan01.com, oauthproxy.com, and the registrations that eventually led us to discover the JASBUG vulnerability in Microsoft products.

Shown below is a selection of domains either registered by JAS on the open market or that are owned by others and were temporarily delegated to JAS in support of this research:

02proxy.com
anams1.com
anams2.com
anams3.com
anams4.com
anams5.com
anams6.com
apa-server2.com
bookchicagolimo.com
bqcproxy.com
bwproxy.com
chicagoetribune.com
chicagohighrise.net
cmsproxy.com
columbusohiozoo.net
corp.com
default-first-site-name.com
etijercorp.com
holasa0.com
holasa01.com
holasa02.com
holasa03.com
ichicagowedding.com
iisproxy.com
lvfs1-2k.com
nalucorp.com
nysproxy.com
oauthproxy.com
ohiomowerdealer.com
oilgasleaseohio.net
pfizercorp.com
prn-03.com
proxy-berlin.com
proxy-chicago.com
proxy-london.com
proxy-singapore.com
rucampinginohio.com
sharnetcorp.com
sipexternal.net
sipinternal.net
taogroupit.com
taolvbes1.com
taolvdc1.com
taolvfile2.com
tiretownchicago.com
vlan01.com
vlan101.com
vlan141.com
vlan142.com
vlan143.com
vlan144.com
vlan145.com
vlan400.com
vlan403.com
vlan404.com
vlanb.com
wjccorp.com
wnadroot.com

Based on our research, knowledge gained by observing DNS namespace collisions in existing TLDs, and a review of research done by others concerning DNS namespace collisions, we classified the behaviors leading to collisions.~~.~~ Organizing the classification into a taxonomy leads to an understanding that: (1)~~ a~~ very few root causes seem to explain the vast majority of colliding behavior, and (2) nearly all root causes appear in all TLDs in differing proportions. Only .corp, .home, and .mail are clear outliers.

The classification was based on: (1) the diversity of querying source IP addresses and Autonomous Systems; (2) the diversity of labels queried; (3) applying sophisticated "randomness detection" to strings and substrings; (4) presence of linguistic terms and colloquialisms in strings and substrings; (5) temporal patterns; and (6) analysis of the Regular Expressions of the labels queried within each TLD and across all TLDs. ~~Aside from improving our understanding of the behavior within .corp, .home, and .mail, we eventually found that the taxonomy does not directly translate to mitigation techniques. Mitigation techniques addressing the small number of root causes are applicable to all TLDs. Dr. Arnoldo Muller-Molina, Founder and Chief Data Scientist of our partner simMachines presented some information about the classification research we performed at the closed DNS-OARC workshop in Warsaw, Poland in May 2014.[47] We also note that the JAS public comment submission included an analysis of colliding queries over a few of the aforementioned dimensions.[48]~~

Additionally, we considered the frequency of occurrence, how easily occurrences were explained by a single etiology, the nature of the string (a long-lived generic term as opposed to a string related to a specific vendor/technology), evidence of use in documentation, examples, and evidence of hardcoded use in software/firmware.

Development of this taxonomy was supported by the data in the Appendices to this report as described in the sections below.

We found that the taxonomy does not directly translate to mitigation techniques but rather mitigation techniques addressing the small number of root causes are applicable to all TLDs.

Because JAS is aware of specific instances where DNS configurations have not yet been remedied, we will list a count of strings in each category as opposed to listing explicit strings, as publishing the discovery of a specific operator's DNS misconfiguration may expose that operator to increased risk. Where possible, JAS has attempted to reach operators privately and provide pointers to ICANN's collision resources.

Below are the categories and counts for the JAS DNS Namespace Collisions Taxonomy. This study attempts to classify the 1,388 unique strings applied-for in the latest ICANN new gTLD round to a set of five basic root causes – the sixth category being unknown for strings that do not fit well into the other five.

~~[47] https://indico.dns-oarc.net//contributionDisplay.py?sessionId=1&contribId=30&confId=19~~

~~[48] http://forum.icann.org/lists/comments-name-collision-05aug13/pdf3WmZlrH3fo.pdf~~

Note this is an attempt to classify the applied-for TLD strings; previous classifications by JAS and others have been attempts to categorize the observed DNS queries.

It's important to note that the precise numbers are certainly debatable and not intended to be individually meaningful; the meaning from this presentation is derived from the categories themselves and the overall high-level proportion between the categories.

Category	Count (of 1388)	%
Likely intentional internal TLD use (name/brand/acronym)	443	32%
Other/unknown	411	30%
Likely ISP/facility suffix	221	16%
Likely intentional internal TLD use (concept/non-brand term)	197	14%
Likely unintentional internal use (other/unknown)	98	7%
Likely unintentional internal use (2LD leakage)	18	1%

5.1 Likely intentional internal TLD use (name/brand/acronym)

This etiology very likely explains more instances of DNS namespace collisions than any other single root cause. This is the scenario where a business or user is intentionally using a TLD internally that is representative of their business or person. If the category described below in 5.3 could be thought of as a "horizontal" pattern, this is the related "vertical" pattern. An example of this scenario would be company NewCo using .newco as the root of their internal DNS.

Less blatant scenarios include situations where hostnames or devices have an internal TLD appended to their configured hostname "printer01.chicago.newco" or are configured into their DNS search path causing the local DNS resolver to generate queries ending in .newco that may leak to the Internet. In speaking with several enterprises impacted in this way, we found that often hostnames are assigned by DHCP and unintended consequences of not well understood DHCP configuration parameters were causing this behavior.

Because of the behavior of modern DNS local stub resolvers, both the mere configuration of a hostname with a TLD extension ("printer01.chicago.newco") and/or intentional configuration of a TLD (".newco") into local or site DNS search path processing may lead to this behavior.[49]

Referring to Appendix A, TLDs falling into this category tend to have a high representation of regular expressions describing hostnames resembling internal machine/device naming schemes ("printer", "server", and the like). If end-user clients receive a hostname from DHCP ending in such a TLD, we also find the presence of "Google Chrome 10 strings" indicating that the Google Chrome browser is being used in the environment (and DNS search path processing is appending the internal TLD to the random string). Referring to Appendix B, TLDs in this category also tend to have lower source diversity.

[49] https://www.icann.org/en/system/files/files/sac-064-en.pdf

5.2 Likely ISP/facility suffix

Scenarios where a local facility or ISP – knowingly or unknowingly – appends a suffix to DNS queries fall into this category. We found a number of examples of hosting providers appending TLDs to what appear to be client DNS queries, likely due to configuration of their recursive resolvers. Similarly, some ISPs use firmware or configurations that cause customer premise equipment (CPE) to append various suffixes to client DNS queries. Several examples of these scenarios have been described publicly as related strings experience controlled interruption.

Like the above scenario, due to the behavior of modern DNS local stub resolvers, both configuration of a hostname with a TLD extension and/or intentional configuration of a TLD into local or site DNS search path processing may lead to this behavior.[50] Another common scenario is an ISP that assigns seemingly "dummy" hostnames to clients using DHCP that are not rooted in a DNS namespace they control.

Referring to Appendix A, TLDs falling into this category tend to exhibit regular expressions that appear to describe client behaviors including: "Google Chrome 10 strings," English words typed into browser URL bars, and DNS label components describing (or even stating outright!) an ISP or service provider. Referring to Appendix B, TLDs in this category also tend to have lower source diversity that correlates to a small number of Autonomous Systems.

The primary difference between 5.1 and 5.2 is the "who:" 5.1 describes the scenario where it appears that the observed behavior is triggered largely by configuration within an enterprise whereas 5.2 describes behaviors where an ISP, vendor, or hosting facility appears to be triggering the observed behavior.

5.3 Likely intentional internal TLD use (concept/non-brand term)

These scenarios involve IT operators internally using TLDs that have some linguistic meaning but are not otherwise tied to their brand. DNS namespace collisions discussed in various technical support fora involving the .prod (for "production") and .app (for "application") namespaces are good examples. This category is the "horizontal" to category 5.1's "vertical." JAS is aware of several large sophisticated technology operations that exhibited this behavior and remedied it following controlled interruption. As with the above categories, DNS search path processing and query devolution is often a factor.

TLDs in this category tend to have higher Appendix B source diversity (as they are being used by multiple entities) and Appendix A regular expressions resembling internal machine/device naming schemes.

[50] ibid.

5.4 Likely unintentional internal use (other/unknown)

JAS applies this category to strings that appear to have limited use related to a single business but otherwise the behavior is not understood and does not fall into any of the above categories. We suspect several of these instances are explained by localized misconfigurations and/or software bugs.

5.5 Likely unintentional internal use (2LD leakage)

This category describes a specific occurrence where an internal TLD appears to be removed from internal DNS queries and the resulting qnames are transmitted to the Internet. For example, a large enterprise using a resource named "files.marketing.chicago.newco" may leak qnames ending in ".chicago" to the Internet if the suffix is removed by DNS search path processing and query devolution. Strictly speaking, this behavior could be considered a subset of one or more of the categories above.

5.6 Other/Unknown and too little data

This category is the catchall for strings either experiencing little or no observable collision activity or the observable behavior was not understood and did not fit into any of the other categories.

5.7 On .corp, .home, and .mail

JAS believes that the widespread use of .corp and .home stems largely from the belief by operators that they may be used safely as private TLDs. RFC 6762, Appendix G, published in 2013,[51] accurately describes this state:

Some network operators setting up private internal networks ("intranets") have used unregistered top-level domains, and some may have used the ".local" top-level domain. [. . .]
We do not recommend use of unregistered top-level domains at all, but should network operators decide to do this, the following top-level domains have been used on private internal networks without the problems caused by trying to reuse ".local" for this purpose:

.intranet.
.internal.
.private.
.corp.
.home.

[51] https://tools.ietf.org/html/rfc6762

.lan.

Based on the data and JAS' experience, "reclaiming" .corp and .home from this unofficial use-case would be a significant undertaking, very likely causing some degree of disruption to a wide range of networks. While not an endorsement of this use-case, RFC 6762's *mere recognition* of the apparent safety of .home and .corp for use in private namespaces is material and seems consistent with the "Convention over Configuration" attitude that permeates so many things on the Internet.

Mike O'Connor generously delegated corp.com to JAS for a period during our analysis. Corp.com may be considered a rough proxy for .corp as on failing a query into .corp many systems will attempt the same query into corp.com in common DNS configurations. The volume of DNS queries into corp.com is nothing short of *staggering*; JAS observed query rates at times exceeding 30 queries/second from a wide range of systems globally. Hundreds of thousands of unique qnames were seen over the relatively short time JAS was delegated corp.com. The rate did not seem to be materially impacted by JAS or Mike O'Connor's publication of multiple controlled interruption-like zones, which may indicate that a number of the sources are devices, non-technically-savvy users, and/or small businesses either unable to understand issues and/or unable to be remediated. Many – but not all – queries seem related to Microsoft Active Directory systems which very often are rooted in ".corp" per an unfortunate Microsoft configuration example more than a decade ago.

The situation with .mail is slightly different. Our analysis of .mail indicates that its use seems less about intentional use rooting an internal namespace and more about legacy configurations. "Reclaiming" .mail from these legacy use-cases would almost certainly be disruptive, although probably not to the extent of .corp and .home.

5.8 Use of Interisle categories in the appendices

JAS used Interisle's label category descriptions when possible to maintain compatibility between reports.[52] Additional categories were added by JAS and appear in the appendices; they are defined below:

Interisle Categories:	
Tables 10-11 (page 47) of the Interisle report[53] define the "INCATx" categories we use in this report. The first row of the Interisle table is INCATB, the second is INCATC, etc. The label 'INCATA' is not used in the JAS reports.	INCATB – INCATL
Additionally, JAS added these categories:	
qname > 253 bytes in length OR sld contains non-LDH characters OR sld byte positions 2-3 are '--' but positions 0-1 are not 'xn' OR sld starts or ends with a hyphen	INVALID
Length of the SLD is < 3 bytes	SHORTSLD
First 4 bytes of sld are 'xn--' but remainder is not valid punycode	INVALIDPUNYCODE
Possibly random 10 character leftmost label based on random detection algorithm	RANDOM10

~~**Sources, methods, and experimental results:** Over the course of this study, we performed a number of experiments and collected a significant amount of data. We talked to a number of vendors, consultants, and end-users experiencing collisions in existing namespaces today. We purchased a number of names and collected data. Several sources contributed data. The types of analysis we performed have been alluded to in this paper and in our presentations at ICANN Singapore, the WPNC in London, and the DNS-OARC workshop in Warsaw. Phase Two of this report will contain the full detail of these activities and relevant datasets.~~

[52] https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf
[53] ibid.