SAC 045
Invalid Top Level Domain Queries at the Root Level of the Domain Name System

A Report from the ICANN
Security and Stability
Advisory Committee
(SSAC)
15 November 2010

This document contains corrections to footnote 2 on page 6 and
Section 8.1 Acknowledgements on page 10

Preface

This is a report by the Security and Stability Advisory Committee (SSAC) on invalid Top Level Domain (TLD) queries at the root level of the domain name system (DNS). The report calls attention to the potential problems that may arise should a new TLD applicant use a string that has been seen with measurable (and meaningful) frequency in a query for resolution by the root system and the root system has previously generated a response.

The SSAC advises the ICANN community and Board on matters relating to the security and integrity of the Internet's naming and address allocation systems. This includes operational matters (e.g., matters pertaining to the correct and reliable operation of the root name system), administrative matters (e.g., matters pertaining to address allocation and Internet number assignment), and registration matters (e.g., matters pertaining to registry and registrar services such as WHOIS). SSAC engages in ongoing threat assessment and risk analysis of the Internet naming and address allocation services to assess where the principal threats to stability and security lie, and advises the ICANN community accordingly. The SSAC has no official authority to regulate, enforce or adjudicate. Those functions belong to others, and the advice offered here should be evaluated on its merits.

The contributors to this report, reference to the committee members' biographies and statements of interest, and committee members' objections to the findings or recommendations in this report, are at end of this report.

This document contains corrections to footnote 2 on page 6 and
Section 8.1 Acknowledgements on page 10

Table of Contents

1. Executive Summary **4**

2. Introduction **4**

3. Background **4**

4. NXDOMAIN Responses From the Root Level of the DNS **5**

5. Finding **7**

6. Recommendations **8**

7. Summary and Conclusions **9**

8. Acknowledgments, Statements of Interests, and Objections and Withdrawals **9**

8.1 Acknowledgments **9**

8.2 Statements of Interest **10**

8.3 Objections and Withdrawals **10**

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

1. Executive Summary

The introduction of new Top Level Domains (TLDs) involves technical considerations of the strings that may be proposed for use by applicants. This report calls attention to the potential problems that may arise should a new TLD applicant use a string that has been seen with measurable (and meaningful) frequency in a query for resolution by the root system and the root system has previously generated a response.

2. Introduction

The introduction of new TLDs involves technical considerations of the strings that may be proposed for use by applicants. With respect to the resolution of TLD strings at the root level of the domain name system (DNS), three conditions exist:

1. The string exists in the root zone and resolves, i.e., a positive result is returned for the query;
2. The string has never been seen in a query for resolution by the root system, i.e., the string has not been delegated and has not been queried;
3. The string has been queried and a root name server has responded to the query with a non-existent domain (NXDOMAIN) result, i.e., the string has not been delegated but has been queried; and
4. The string was resolved by root name servers at one time in the past but has been removed from the root zone, i.e., the string is a previously delegated string, and root name servers have returned positive responses to queries for that string.

This report calls attention to conditions (2) and (3) above and, specifically, the potential problems that may arise should a new TLD applicant use a string that has been seen with measurable (and meaningful) frequency in a query for resolution by the root system and the root system has previously generated a response.

3. Background

In the normal course of domain name resolution, a client on a host or application will query a resolver for resource records associated with a domain name. If the resolver can provide an answer to the query from local (cached) information, it does so. If the resolver cannot provide an answer, it uses a recursive process to resolve the domain name. Specifically, the resolver queries a root name server for the resource records associated with a domain name. For example, if the domain name in the query were `www.example.com`, the resolver asks a root name server for the full name. However, since the root servers do not maintain information about the full name, a referral response is returned that contains the list of name servers for the `.COM` TLD. The resolver next asks one of the `.COM` TLD name servers for the resource records for the full name. Since the .COM TLD name servers do not have the answer, a referral response is returned that contains the list of name servers for the `example.com domain.` The resolver then

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

asks one of `example.com`'s name servers for the resource records associated with `www.example.com`, and the answer is returned. If the TLD is unknown to the root servers (i.e., if the rightmost string in the domain name before the (optional) terminating "." is not in the root zone file), the resolver receives a negative response (`NXDOMAIN`), which is then relayed back to the requesting client.

4. NXDOMAIN Responses From the Root Level of the DNS

According to analyses of data collected by the Domain Name System Operations, Analysis, and Research Center (DNS-OARC) and reported by the Day in The Life of the Internet (DITL) project certain strings repeatedly appear at the root level of the DNS in queries seeking to resolve top-level domain (TLD) labels.[1] Such strings have not been delegated (included in the root zone). Figure 1 depicts the top invalid TLDs:

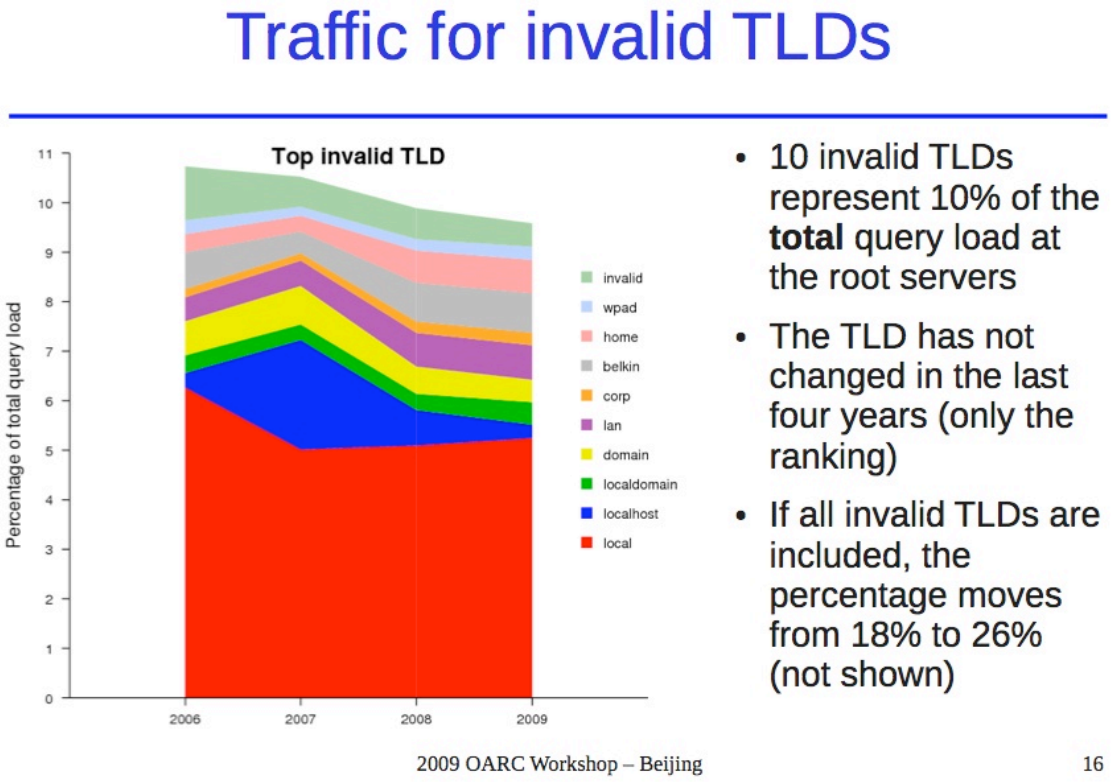

Figure 1. Traffic for invalid TLDs (2009)

These queries are wrongly directed at root name servers as a result of configuration errors or incorrect invocation of DNS in configurations where name spaces other than the DNS

[1] See "DNS Research Update from CAIDA: Status and Recent Experiences," KC Claffy, Root Server System Advisory Committee (RSSAC), March 22, 2009 <http://www.caida.org/publications/presentations/**2009**/rssac_dns/rssac_dns.pdf> and "DITL 2009: Analysis and Results of Four Years of DITL," Sebastian Castro, New Zealand Registry Services and The Cooperative Association for Internet Data Analysis (CAIDA), 2009 OARC Workshop, Beijing, China, <http://www.dns-oarc.net/files/workshop-**2009**11/Sebastian_Castro.pdf>.

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

(e.g., Microsoft's "WINS") are used on private networks. (This "leakage" is part of a broader set of invalid queries referred to as DNS pollution.) From these data, we note the following:

1. Currently, root name servers return NXDOMAIN responses to queries containing a variety of strings. According to a CAIDA report[1], NXDOMAIN responses (measured over a 24 hour period) account for more than 25 percent of the total responses from root name servers observed in the study, and the top ten such strings account for 10 percent of the total query load at the observed root name servers.[2]
2. In the future, a new TLD applicant could apply for a string that has appeared at the root. If the application (and string) were to be approved and the TLD included in the root zone, queries to the root level of the DNS for a string that hitherto returned NXDOMAIN would begin to return positive responses containing name servers of the new TLD.
3. It is likely that many of the same conditions that cause the current set of invalid TLD queries to appear at the root level of the DNS will persist despite efforts to encourage end users, private networks, software and equipment manufacturers to correct configuration and programming errors.
4. The behavior of the DNS for some end users and private networks will therefore be altered. In particular, the change from a NXDOMAIN response to a positive response will result in resolvers continuing recursive resolution. Consider what would happen if the string `.lan` (one of the top 10 invalid TLDs) were to be approved as a new TLD label. Currently, a client that queries `www.example.lan` receives an NXDOMAIN response from a root name server. Once the `.lan` TLD is approved and instantiated in the root zone, that client will begin to receive referral responses for the same query containing resource records for the `.lan` TLD name servers, and will query one of `.lan`'s name servers with the original query. If `example.lan` is registered in the `.lan TLD`, the resolver continues recursion and queries `example.lan` for the web site, which now resolves to a public server what the private network operator using `.lan` expected to be a local name for a local server.
5. The TLD registry operator for `.lan` will "inherit" query traffic. Whereas the root system is provisioned with sufficient capacity to manage the volume of all invalid TLDs without incident, the `.lan` TLD registry operator may not be prepared to deal with tens of millions of hitherto invalid queries

[2] The most frequently observed invalid TLDs in the sampled data from observed root name servers during 2006-2009 include strings such as `local, localhost, lan, home, domain, localdomain, corp`, and `belkin`. From these data, `local` appears to be the most frequent cause of a negative response. Other strings that appear with frequency include `wpad`, `corp`, `maps`, `html`, `router`, `host`, `mshome,` and `htm`. From George Kirikos, "Most Popular Invalid TLDs Should Be Reserved," June 18, 2009 <http://www.circleid.com/posts/20090618_most_popular_invalid_tlds_should_be_reserved/>

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

6. The `.lan` TLD registry operator – and generally, any TLD registry operator that chooses a string that has been queried with meaningful frequency at the root – potentially inherits millions of queries per day. These queries represent data that can be mined or utilized by the registry operator.

The introduction of new TLDs creates the potential for another "inheritance" condition. The scenario is similar in some respects to a re-delegation or decommissioning of a country code TLD. [3] In this scenario the following actions may occur:

- ICANN approves a new TLD registry and string. The string is delegated, included in the root zone, and the root system returns referrals for this new string.
- The registry ceases operation and after a period of time, the TLD is decommissioned and the delegation is removed from the root zone. The root system returns NXDOMAIN responses for this TLD string.
- ICANN accepts an application for a new TLD that intends to use the decommissioned string. (Note that the ICANN New gTLD Draft Applicant Guidebook is silent on this at the moment.) The string is delegated a second time, included in the root zone, and the root system returns referrals for this new string.
- The new registry inherits queries for domains registered under the old version of the same string. Certain of the labels registered under the original TLD registry may persist in Uniform Resource Locators (URLs). If Resource Record Sets (RRsets) of these domains were DNS Security Extensions (DNSSEC)-signed then DNSSEC-aware clients would be able to note the change; however, other clients would accept the referral under circumstances where NXDOMAIN might have been more appropriate.

5. Finding

This report presents the following finding:

Finding: ICANN should make applicants for new TLDs aware of the following: Any new TLD registry operator may experience unanticipated queries and some TLDs may experience a non-trivial load of unanticipated queries if the label it chooses corresponds to TLDs that have historically seen queries.

Studies[1] illustrate that the amount of inherited query traffic could be considerable, i.e., on the order of millions of queries per day, should the applicant's chosen string be one that appears frequently at the root. While millions of queries per day is manageable from an

[3] See Letter from Leslie Daigle, Chair, Internet Architecture Board to Oliver Smoot, President, International Organization for Standardization (ISO), September 26, 2003, <http://www.iab.org/documents/correspondence/2003-09-25-iso-cs-code.html> and Letter from Leslie Daigle, Chair, Internet Architecture Board to Paul Twomey, President and Chief Executive Officer, Internet Corporation for Assigned Names and Numbers (ICANN), <http://www.iab.org/documents/correspondence/2003-09-25-icann-cs-code.html>.

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

operational perspective, it is prudent for ICANN to make applicants aware of the potential for inherited traffic so they are prepared to manage the volume, and will thus minimize the possibility of operational difficulties that would pose a stability or availability problem for their registrants and users.

In addition, parties other than the TLD applicant may be affected, including parties whose systems are currently generating invalid TLD queries and registrants of domains in the TLD. Specifically, parties generating invalid TLD queries and receiving NXDOMAIN from the root servers today will now receive referrals. Whereas the NXDOMAIN forces the querying application or user into an error resolution condition, the referral response from a root name server could cause recursion to continue (consider again the `.lan` scenario describe above), with unpredictable results for the user.

6. Recommendations

This report presents the following recommendations:

Recommendation (1): The SSAC recommends that ICANN promote a general awareness of the potential problems that may occur when a query for a TLD string that has historically resulted in a negative response begins to resolve to a new TLD. Specifically, ICANN should:

- Study invalid TLD query data at the root level of the DNS and contact hardware and software vendors to fix any programming errors that might have resulted in those invalid TLD queries. The SSAC is currently exploring one such problem as a case study, and the vendor is reviewing its software. Future efforts to contact hardware or software vendors, however, are outside SSAC's remit. ICANN should consider what if any organization is better suited to continue this activity.
- Contact organizations that are associated with strings that are frequently queried at the root. Forewarn organizations who send many invalid queries for TLDs that are about to become valid, so they may mitigate or eliminate such queries before they induce referrals rather than NXDOMAIN responses from root servers.
- Educate users so that, eventually, private networks and individual hosts do not attempt to resolve local names via the root system of the public DNS.

Recommendation (2): The SSAC recommends that ICANN consider the following in the context of the new gTLD program.

- Prohibit the delegation of certain TLD strings. RFC 2606, "Reserved Top Level Domain Names," currently prohibits a list of strings, including `test`, `example`, `invalid`, and `localhost`.[4] ICANN should coordinate with the community to

[4] See RFC 2606, "Reserved Top Level Domain Names," <http://www.faqs.org/rfcs/rfc2606.html>.

This document contains corrections to footnote 2 on page 6 and
Section 8.1 Acknowledgements on page 10

identify a more complete set of principles than the amount of traffic observed at the root as invalid queries as the basis for prohibiting the delegation of additional strings to those already identified in RFC 2606.

- Alert the applicant during the string evaluation process about the pre-existence of invalid TLD queries to the applicant's string. ICANN should coordinate with the community to identify a threshold of traffic observed at the root as the basis for such notification.
- Define circumstances where a previously delegated string may be re-used, or prohibit the practice.

7. Summary and Conclusions

In this report, we call attention to the potential problems that may arise should a new TLD applicant use a string that has been seen with measurable (and meaningful) frequency in a query for resolution by the root system and the root system has previously generated a response. We find that any new TLD registry operator may experience unanticipated queries and that some TLDs may experience a non-trivial load of unanticipated queries if the label it chooses corresponds to TLDs that have historically seen queries. We recommend that ICANN inform new TLD applicants of the problems that can arise when a previously seen string is added to the root zone as a TLD label and that ICANN should coordinate with the community to identify principles that can serve as the basis for prohibiting the delegation of strings that may introduce security or stability problems at the root level of the DNS.

8. Acknowledgments, Statements of Interests, and Objections and Withdrawals

In the interest of greater transparency, we have added these sections to our documents to provide the reader information on three aspects of our process. The Acknowledgments section lists the members who contributed to this particular document. The Biographies and Statements of Interest section points to the biographies of the Committee members and any conflicts of interest, real, apparent or potential, that may bear on the material in this document. The Objections and Withdrawals section provides a place for individual members to disagree with the content of this document or the process for preparing it.

8.1 Acknowledgments

The committee wishes to thank the following SSAC members and invited guests for their time, contributions, and review in producing this Report.

Sebastian Castro
KC Claffy
David Conrad
Stephen Crocker

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10

Invalid Top Level Domain Queries at the Root Level of the Domain Name System

George Kirikos
Ray Plzak
Ram Mohan

8.2 Statements of Interest

SSAC member biographical information and Statements of Interest are available at: http://www.icann.org/en/committees/security/biographies-07jul10-en.htm.

8.3 Objections and Withdrawals

There are no objections or withdrawals.

This document contains corrections to footnote 2 on page 6 and Section 8.1 Acknowledgements on page 10