

SAC052

SSAC Advisory on Delegation of Single-Character Internationalized Domain Name Top-Level Domains



An Advisory from the ICANN
Security and Stability
Advisory Committee
(SSAC)
31 January 2012

Preface

This is an Advisory of the Security and Stability Advisory Committee (SSAC). 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). The 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 Advisory, reference to the committee members' biographies and statements of interest, and committee members' objections to the findings or recommendations in this report are included at end of this Advisory.

Table of Contents

1. Executive Summary	4
2. Introduction	5
3. Findings	5
4. Recommendations.....	8
5. Acknowledgments, Statements of Interests, and Objections, and Withdrawals.....	10
5.1 Acknowledgments	10
5.2 Statements of Interest.....	11
5.3 Objections and Withdrawals.....	11
6. References.....	11

1. Executive Summary

At the request of the ICANN Board, the SSAC provides its advice on the *security & stability* aspects of delegating single-character internationalized domain name (IDN) top-level domains.¹

Because IDN U-labels are converted into American Standard Code for Information Interchange (ASCII) A-labels before they are processed directly by the domain name system (DNS), the SSAC does not see significant *security* concerns with the *delegation* of single-character IDN TLDs.² However, the risk of *user confusion* is higher for single-character TLDs than for TLDs with more than one character. It is also evident that review and modifications to the String Similarity Review are required; in parallel, much work remains to be completed on variant handling and management.

The SSAC makes two recommendations:

1. Given the potential for user confusion and the currently unfinished work on string similarity and IDN variants, the SSAC recommends a very conservative approach to the delegation of single-character IDN top-level domains. In particular, ICANN should disallow by default the delegation of all single-character IDN TLDs in all scripts; exceptions are possible, but only after careful consideration of each individual case.
2. Because important relevant work on string similarity, IDN variant issues, and TLD label syntax is currently underway within ICANN, the Internet Engineering Task Force (IETF), and other bodies, ICANN should review the Findings of this report, and any policies that it adopts in response to recommendations made in this document, no later than one year after the three work items mentioned above have been completed.

¹ICANN Board of Directors Resolutions on Single Character IDN Update
<<http://www.icann.org/en/minutes/resolutions-25aug11-en.htm#5>>.

²A domain name consists of a series of "labels" (separated by "dots"). The ASCII form of an IDN label is termed an "A-label." An A-label conforms to the Letter-Digit-Hyphen (LDH) constraint on labels as defined by the DNS standards. All operations defined in the DNS protocol use A-labels exclusively. The Unicode form, which a user expects to be displayed, is termed a "U-label." A special form of "ASCII compatible encoding" (ACE) is applied to a U-label to produce a corresponding A-label. The transformation is symmetric: one can derive a U-label from an A-label for the purpose of displaying the domain name using characters from a local script so that a user sees a familiar script rather than a less recognizable A-label.

2. Introduction

The Internet's domain name system provides a distributed lookup mechanism for hierarchically allocated strings called domain names. It allows users to refer to web sites and other resources using easier-to-remember domain names (such as "www.icann.org") rather than the all-numeric Internet Protocol (IP) addresses (such as "192.0.32.7") assigned to computers on the Internet. Each domain name consists of a sequence of character strings (called "labels"). In most written forms the labels in a domain name are separated by dots, and the right-most label (or left-most for scripts that are written right-to-left) in a domain name is referred to as its "top-level domain" (TLD).

Currently, each TLD label consists either of (a) two or more alphabetic characters (a-z) from the ASCII character set, or (b) two or more IDN characters (for IDN country code TLDs), rendered via a Punycode encoding mechanism into an ASCII letter, digit, or hyphen (LDH) form for processing by the DNS. Recently, the Joint country code Names Supporting Organization (ccNSO)-Generic Names Supporting Organization (GNSO) IDN Working Group (JIG) recommended that single-character TLDs be accepted in the programs related to IDN ccTLDs and New gTLDs.³

This advisory is the SSAC's response to the ICANN Board's request for advice concerning the security and stability impact of delegating single-character top-level domains. It has two parts: Section 3 summarizes three key findings related to the delegation of single-character IDN TLDs, and Section 4 offers specific recommendations for the Board to consider.

3. Findings

What is a "single character"?

The term "single character" is easier to define for some scripts than for others. In particular, it does not correspond to "one Unicode code point," as glyphs that would be recognized by users as "single characters" can arise from sequences of one or more Unicode code points. As the context here is the potential for user confusion, the SSAC considers the term "single character" to mean one character in a U-label, as it would be recognized by a user familiar with the script of the U-label, represented by one or more Unicode code points. As stored in and processed by the DNS, a single-character U-label would be represented in the corresponding A-label as a minimum of seven ASCII characters.

³ICANN Joint ccNSO-GNSO IDN Working Group (JIG) Final Report on Single Character IDN TLDs (30 March 2011) <<http://gns0.icann.org/drafts/jig-final-report-30mar11-en.pdf>>.

SSAC Advisory on the Delegation of Single-Character Internationalized Domain Name Top-Level Domains

The SSAC notes that the syntax for allowed TLD labels in the DNS is not clearly applicable to the encoding of IDNs. The IETF is working on a concise specification of the TLD label syntax based on existing syntax documentation, extended minimally to accommodate IDNs.⁴ The outcome of this work is expected to define what subset of allowed Unicode code points defined in RFC 5892 would be available for use in TLD labels.⁵ This is an important consideration to the delegation of IDN TLDs of any length, and of single-character IDN TLDs specifically with reference to M-class Unicode code points, which match a variety of non-spacing and combining marks.

Finding 1: Single-character TLDs are more likely to cause user confusion than TLDs with more than one character.

In general, the more characters a domain label has, the easier it is for a user to "infer the context." For example, if one character in a multi-character label is confusingly similar to another character not in the label, so long as the other characters are not confusing, the user can infer the context based on the non-confusing characters in the label. In the case of two-character TLDs, a special classification algorithm has been developed for the IDN ccTLD Fast Track program that applies to the assessment of confusability with the ISO 3166-1 alpha-2 country codes.⁶ This classification algorithm specifically accepts confusability regarding one character if the other character is distinct. This algorithm recognizes that in general the presence of one or more non-confusable characters in a label mitigates the confusability of one or more other characters in the same label.

Further, code points in closely related script blocks may represent characters that are confusingly similar despite belonging to different scripts. For example, Latin Small Letter A (U+0061), Greek Small Letter Alpha (U+03B1), and Cyrillic Small Letter A (U+0430) are visually identical in some fonts. As a result, single-character TLDs expressed in Latin, Greek, or Cyrillic are more likely to cause user confusion. Similar examples may be found in other script-groups, including Gurmukhi/Bangla/Devanagari in South Asia and Thai/Lao in South East Asia.⁷

⁴L. J. Liman, and Joe Abley, "Top Level Domain Name Specification", IETF Work in Progress (draft-liman-tld-names-06), (2011) <<http://tools.ietf.org/html/draft-liman-tld-names>>.

⁵Patrik Fältström, "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892 (August 2010) <<http://tools.ietf.org/html/rfc5892>>.

⁶Tina Dam, "Clearing the Confusion (Fast Track)" (2012) <<http://blog.icann.org/2010/03/clearing-the-confusion-fast-track/>>.

⁷See Section 4.1 and Table 7 in Devanagari Case Study Team report for IDN Variant Issues Project at for a more comprehensive list. <<http://www.icann.org/en/topics/new-gtlds/devanagari-vip-issues-report-03oct11-en.pdf>>.

SSAC Advisory on the Delegation of Single-Character Internationalized Domain Name Top-Level Domains

The SSAC is not aware of a comprehensive inventory of confusable scripts, nor are we aware of current work to identify groups of scripts that are related in such a way that they present an intrinsic risk of inter-script user confusion. A responsibly conservative approach is therefore to assume the likelihood that there are other “undiscovered” groups of intrinsically confusable scripts.

For ideographic scripts such as Han, not only can a single character represent a complete “word” or idea, but in some cases different single characters can represent the same “word” or idea. Were ICANN to delegate each such different single character as a TLD label (whether to the same or to a different registrant), users would likely be subject to confusion based on varying deployments of the single character, defined by registry policy. The problem of “synchronization” of TLDs previously has been studied and it is clear that there are no unified technical approaches that work consistently in the DNS.⁸ To the extent that two or more different single characters that have the same meaning (variants) may be delegated, a context-free single-character TLD could represent a *higher* degree of confusability than an equivalent multi-character TLD with at least one non-confusable character.

Finding 2: No other significant security concerns are apparent with the delegation of single-character TLDs.

In the DNS and in applications, IDN labels are either U-labels (a sequence of Unicode code points) or A-labels (a sequence of ASCII characters). A U-label is transformed into an A-label using Punycode, an encoding mechanism specified by the IDNA protocol. The DNS sees only the A-label form of IDNs. Therefore, as far as the DNS is concerned, all domain names are just ASCII and the DNS resolves all such names in a uniform and predictable manner. In this sense the introduction of IDNs at the top level of the DNS does not change anything about the way in which the DNS itself operates, although it has many consequences for the operation of applications that handle IDNs as U-labels.

Finding 3: Current work on string similarity and variant issues has not been completed.

ICANN is working on updating the string similarity review process. Recently ICANN published an updated version of the IDN ccTLD Fast Track Process, which describes conditions for the delegation of IDN ccTLD labels that were previously disallowed.⁹

⁸Internet Corporation for Assigned Names and Numbers (ICANN) (2011b), A Study of Issues Related to the Management of IDN Variant TLDs. Marina Del Rey, CA (23 December 2011) <<http://www.icann.org/en/topics/new-gtlds/idn-vip-integrated-issues-23dec11-en.pdf>>.

⁹Internet Corporation for Assigned Names and Numbers (ICANN) (2011a), Guidelines for the Implementation of Internationalized Domain Names Version 3.0 (2011a). Marina Del Rey, CA (02 September 2011) <<http://www.icann.org/en/topics/idn/idn-guidelines-02sep11-en.htm>>.

SSAC Advisory on the Delegation of Single-Character Internationalized Domain Name Top-Level Domains

Further review and modifications of the String Similarity Review process are clearly required to reduce ambiguity and increase consistency in the process.

The IDN Variant Issues Project (VIP) has also just published its first draft report.¹⁰ In the DNS environment today, there is no accepted definition for what may constitute a “variant” relationship among top-level labels, nor is there a uniform “variant management” mechanism for the top level, although solutions to pieces of the problem have been proposed. Several communities have indicated an urgent need for solutions in this area, to support deployment of the full range of products and services made possible by bringing IDN capabilities to the namespace.

The VIP report reveals a tension between the unmistakable interest in creating greater functionality to address a range of potential variant cases and the difficulty of using the DNS to meet these objectives. In proposing a classification of “variant” cases and outlining what cases appear most appropriate for development of solutions, the report attempts to highlight the cost-benefit analysis that needs to be done before undertaking any implementation. This important community-led work and its findings will lead to further changes in the implementation of IDNs at the top level.

It is clear to the SSAC that further work is required in both the string similarity process area and the variant issues area in order to further understand and promote uniformity of access to IDN TLDs on the Internet.

4. Recommendations

Recommendation 1: Given the potential for user confusion and the currently unfinished work on string similarity and IDN variants, the SSAC recommends a very conservative approach to the delegation of single-character IDN top-level domains.

In particular, until ICANN completes its work on user confusion/string similarity and IDN variants, the SSAC recommends:

1. Delegation of all single-character IDN TLDs in all scripts should be disallowed by default.
2. Exceptions may be made for some scripts, but only after careful consideration of potential confusability both within and across scripts. Such consideration should invite comments from the technical and linguistic community, and from ICANN’s advisory committees.

¹⁰Internet Corporation for Assigned Names and Numbers (ICANN) (2011b), A Study of Issues Related to the Management of IDN Variant TLDs. Marina Del Rey, CA (23 December 2011) <<http://www.icann.org/en/topics/new-gtlds/idn-vip-integrated-issues-23dec11-en.pdf>>.

SSAC Advisory on the Delegation of Single-Character Internationalized Domain Name Top-Level Domains

3. Single-character TLD applications in an exceptionally allowed script should be accepted only when there is clear evidence that there is no risk of user confusion. Each applied-for single-character TLD label must be explicitly examined across scripts to ensure that there is absolutely no possibility of user confusion within or across scripts.
4. ICANN should consult with the technical and linguistic community to determine which scripts, if any, should be restricted with respect to the delegation of single-character TLDs, and how any such restrictions should be defined, and how such restrictions may be relaxed if appropriate.
5. ICANN should take into consideration the outcome of the IETF work on the creation of a concise specification of the TLD label syntax based on existing syntax documentation, extended minimally to accommodate IDNs.¹¹
6. ICANN should consider adopting the following guidelines regarding its consideration of which scripts and code points could be accepted as exceptions:
 - a) The code point must be PVALID according to IDNA2008.
 - b) The code point is from one of the following Unicode categories: lower case letter (Ll), upper case letter (Lu), and other letter (Lo) as defined by the Unicode Standard.¹²
 - c) Some single-character IDN TLDs are composed of multiple Unicode code points, which may include non Lx-class code points. These should be subjected to a more stringent technical and confusability analysis, whose criteria should be well defined and made public.
 - d) The script in which an exception is made and a single character IDN is allowed should not have characters that are intrinsically confusable with characters of another script (for example, Latin/Greek/Cyrillic, Lao/Thai, etc.).
 - e) The existing and extended rules of confusability must be met. Single-character code points must explicitly be examined across scripts. Denial of a single-character TLD application does not imply blocking of the script. Similarly,

¹¹L. J. Liman and Joe Abley, “Top Level Domain Name Specification”, IETF Work in Progress (draft-liman-tld-names-06), (2011) <<http://tools.ietf.org/html/draft-liman-tld-names>>.

¹²The Unicode Consortium, “The Unicode Standard, Version 6.0”, (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6): <<http://www.unicode.org/versions/Unicode6.0.0/>>.

acceptance of a single-character TLD application does not imply acceptance of the script.

- f) If a script is allowed, a distinct and explicit specification of which subset of the script is available for single-character TLDs should be required prior to the acceptance of a single-character TLD application. By default all characters are disallowed, even when a script is allowed, and an explicit single-character-TLD-allowed list must be generated for each case.

Recommendation 2: Because important relevant work on string similarity, IDN variant issues, and TLD label syntax is currently underway within ICANN, the IETF, and other bodies, ICANN should review the Findings of this report, and any policies that it adopts in response to Recommendation 1, no later than one year after the three work items mentioned above have been completed.

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

In the interest of greater transparency, these sections provide the reader information on three aspects of our process. The Acknowledgments section lists the members who contributed to this particular document. The 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.

5.1 Acknowledgments

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

SSAC Members

Bruce Tonkin
Jim Galvin
Lyman Chapin
Patrik Fältström
Ram Mohan
Sarmad Hussain

ICANN staff

Steve Sheng (editor)
Dave Piscitello
Xiaodong Lee

5.2 Statements of Interest

SSAC member biographical information and Statements of Interest are available at: <http://www.icann.org/en/committees/security/biographies-31jan12-en.htm>.

5.3 Objections and Withdrawals

There were no objections or withdrawals

6. References

1. Dam, T. (2010), "Clearing the Confusion (Fast Track)", Available at <<http://blog.icann.org/2010/03/clearing-the-confusion-fast-track/>>.
2. Faltstrom, P., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892, August 2010.
3. Internet Corporation for Assigned Names and Numbers (ICANN) (2011a), Guidelines for the Implementation of Internationalized Domain Names Version 3.0 (2011a). Marina Del Rey, CA: ICANN. Available at: <<http://www.icann.org/en/topics/idn/idn-guidelines-02sep11-en.htm>>.
4. Internet Corporation for Assigned Names and Numbers (ICANN) (2011b), A Study of Issues Related to the Management of IDN Variant TLDs. Marina Del Rey, CA: ICANN. Retrieved January 10, 2012, from <<http://www.icann.org/en/topics/new-gtlds/idn-vip-integrated-issues-23dec11-en.pdf>>.
5. ICANN Joint ccNSO-GNSO IDN Working Group (JIG) (2011). Final Report on Single Character IDN TLDs. Available at: <<http://gns0.icann.org/drafts/jig-final-report-30mar11-en.pdf>>.
6. ICANN Board of Directors (ICANN Board) (2011). Resolutions on Single Character IDN Update. Available at: <<http://www.icann.org/en/minutes/resolutions-25aug11-en.htm#5>>.
7. Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, August 2010.

SSAC Advisory on the Delegation of Single-Character Internationalized Domain Name Top-Level Domains

8. Liman, L. J. and Abley, J., “Top Level Domain Name Specification”, IETF Work in Progress (draft-liman-tld-names-06), 2011. Available at <http://tools.ietf.org/html/draft-liman-tld-names>.
9. The Unicode Consortium, “The Unicode Standard, Version 6.0”, (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6). <http://www.unicode.org/versions/Unicode6.0.0/>.