Abstract

The syntax for allowed Top-Level Domain (TLD) labels in the Domain Name System (DNS) is not clearly applicable to the encoding of Internationalised Domain Names (IDNs) as TLDs.

This document provides a concise specification of TLD label syntax based on existing syntax documentation, extended minimally to accommodate IDNs.

This document updates RFC1123.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on June 1, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents.
carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction .............................................. 3
2. Definitions .............................................. 4
3. Background ............................................... 5
4. TLD Label Syntax Specification .......................... 7
5. Policy Considerations ..................................... 8
6. IANA Considerations ....................................... 9
7. Security Considerations ................................... 10
8. Acknowledgements ......................................... 11
9. References ................................................ 12
   9.1. Normative References ................................. 12
   9.2. Informative References ............................... 12
Appendix A. Change History .................................. 13
   A.1. draft-liman-tld-names-06 ............................ 13
   A.2. draft-liman-tld-names-05 ............................ 13
   A.3. draft-liman-tld-names-04 ............................ 13
   A.4. draft-liman-tld-names-03 ............................ 13
   A.5. draft-liman-tld-names-02 ............................ 13
   A.6. draft-liman-tld-names-01 ............................ 13
   A.7. draft-liman-tld-names-00 ............................ 13
   A.8. Version Identification Tag .......................... 14
Authors’ Addresses ......................................... 15
1. Introduction

The syntax of TLD labels ("TLD DNS-Labels", as defined in Section 2) is specified in [RFC1123], where such labels are asserted to be "alphabetic" within a section of that document entitled "DISCUSSION". This can be interpreted as requiring that the hyphen character ("-") and numeric digits be excluded from TLD DNS-Labels. Such a restriction would not accommodate the US-ASCII encoding of Internationalised Domain Names (IDNs), as specified in [RFC5890]. A more detailed discussion of the existing specifications can be found in Section 3.

This document extends the syntax of allowable TLD DNS-Labels to support IDNs, but places some restrictions on the choice of IDN labels. These restrictions are intended to be consistent with the existing specification for US-ASCII TLD DNS-Labels. See Section 4 for the updated specification.

This document focuses narrowly on the issue of allowable DNS-Labels in TLDs and does not (and is not intended to) make any other changes or clarifications to existing domain name syntax rules.

It is carefully noted that the specification in this document is not the only factor in choosing suitable TLD DNS-Labels, and that many considerations external to the IETF are included in that wider policy. See Section 5 for more discussion of policy considerations.
2. Definitions

The term DNS-Label is used in this document to have precisely the same meaning as the term "label", as introduced in [RFC1034], section 3.1. A DNS-Label denotes one node in a DNS tree. A DNS-Label is zero to 63 octets in length. The term "DNS-Label" refers exclusively to the "wire format" of the label, and not to any presentation format of the label.

A Top-Level Domain (TLD) DNS-Label is the right-most ("highest-level") DNS-Label in a fully-qualified domain name.

The terms A-Label and U-Label are used in this document as defined in [RFC5890].
3. Background

[RFC0952] defines a host name as follows:

‘A "name" ... is a text string up to 24 characters drawn from the alphabet (A-Z), digits (0-9), minus sign (-), and period (.). Note that periods are only allowed when they serve to delimit components of "domain style names". (See RFC-921, "Domain Name System Implementation Schedule", for background). No blank or space characters are permitted as part of a name. No distinction is made between upper and lower case. The first character must be an alpha character. The last character must not be a minus sign or period.’  [Unnumbered section titled "ASSUMPTIONS", first paragraph]

[RFC1123] reaffirms this definition, but makes one change to the syntax:

‘The syntax of a legal Internet host name was specified in RFC-952 [DNS:4]. One aspect of host name syntax is hereby changed: the restriction on the first character is relaxed to allow either a letter or a digit. Host software MUST support this more liberal syntax.’  [Section 2.1]

In addition, the DISCUSSION section of Section 2.1 says:

‘However, a valid host name can never have the dotted-decimal form #.#.#.#, since at least the highest-level component label will be alphabetic.’  [Section 2.1]

Some implementers may have understood the above phrase ‘will be alphabetic’ to be a protocol restriction.

Neither [RFC0952] nor [RFC1123] explicitly states the reasons for these restrictions. It might be supposed that human factors were a consideration; [RFC1123] appears to suggest that one of the reasons was to prevent confusion between dotted-decimal IPv4 addresses and host domain names. In any case, it is reasonable to believe that the restrictions have been assumed in some deployed software, and that changes to the rules should be undertaken with caution.

The Internationalised Domain Names in Applications 2008 specification (IDNA2008) [RFC5891] [RFC5892] provides a protocol for encoding Unicode strings in DNS-Labels. The Unicode string used by applications is known as a U-Label; its corresponding encoding in the DNS is known as an A-Label. The terms A-Label and U-Label are used in this document as defined in [RFC5890]. Valid A-Labels always contain non-alphabetic characters.
In order to accommodate the wish to express TLD names in scripts other than Latin (or rather, the US-ASCII subset of Latin), it is necessary to allow non-alphabetic characters in the corresponding TLD DNS-Labels. To minimize changes, the U-label form of a TLD name is restricted in ways functionally compatible with the restrictions (from [RFC0952] and [RFC1123]) on US-ASCII TLD names, by applying rules analogous to those already imposed on US-ASCII TLD DNS-Labels to TLD U-labels.

However, deployed software that checks DNS top-level labels for conformance with an alphabetic restriction will not recognize such corresponding A-Labels (i.e., U-labels represented in their US-ASCII form).
4. TLD Label Syntax Specification

This document relaxes the existing specification to allow TLD DNS-Labels to be well-formed A-Labels, but places restrictions on their corresponding U-Labels. That is, not every well-formed A-Label is a valid TLD DNS-Label.

The ABNF expression that matches a valid TLD DNS-Label is as follows:

\[
\text{tld-dns-label} = \text{traditional-tld-label} / \text{idn-label}
\]

\[
\text{traditional-tld-label} = 1^63(\text{ALPHA})
\]

\[
\text{idn-label} = \text{Restricted-A-Label}
\]

\[
\text{ALPHA} = \%x41-5A / \%x61-7A ; A-Z / a-z
\]

A Restricted-A-Label is a DNS-Label which satisfies all the following conditions:

1. the DNS-Label is a valid A-Label according to [RFC5890];

2. the derived property value of all code points, as defined by [RFC5890], is PVALID;

3. the general category of all code points, is one of \{ Ll, Lo, Lm, Mn, Mc \}.

This new specification reflects current practice in registration of TLD names by the IANA, extended to accommodate IDNs.
5. Policy Considerations

This document provides a technical specification that limits the set of TLD DNS-Labels that are available for assignment; it does not aim to encapsulate the full policy framework within which TLD names are chosen.

At the time of writing, the policy under which TLD names are chosen is developed and maintained by ICANN in consultation with a wide base of stakeholders. As the Internet continues to grow to serve new user communities, applications and services, it is to be expected that the corresponding policy will be changed accordingly.
6. IANA Considerations

While this document makes no requests of the IANA, management of the root zone is an IANA function. This document expands the set of strings permitted for delegation from the root zone, and hence establishes new limits for the corresponding IANA policy.
7. Security Considerations

This document is believed to have limited security implications.

General discussion about the security effects of internationalized labels can be found in [RFC5890], section 4. Those considerations apply equally to TLD labels.

The creation of new TLDs has the potential to conflict with software which (for example) predates and correspondingly does not accommodate new TLD names. Such software problems might in turn lead to security vulnerabilities, e.g. in the case where a DNS name specified by a user is truncated or otherwise misinterpreted, causing an application to interact with a different remote host from that which the user intended. It should be noted that this is not a new phenomenon, and has been observed following the creation of new (US-ASCII) TLD names prior to the publication of this document.

The issue that some Unicode characters can be confused with each other is discussed at length in the Security Considerations section of [RFC5890].
8. Acknowledgements

Tina Dam, Patrik Faltstrom, John Klensin, Thomas Narten and Andrew Sullivan contributed text to this document, and their contributions are hereby acknowledged.
9. References

9.1. Normative References


9.2. Informative References


Appendix A. Change History

This section (and sub-sections) should be removed before publication.

A.1. draft-liman-tld-names-06

Add Mc as an allowable code-point, required for names in Devanagari script.

A.2. draft-liman-tld-names-05

New affiliation and address for Liman, due to company merger.

A.3. draft-liman-tld-names-04

Removed subjective and unverified statements regarding deployed software. Replaced with more generic text. Polishing a few expressions to make them less obtrusive. Removed confusing paragraph after ABNF table. Updated some references that are now published as RFCs.

A.4. draft-liman-tld-names-03

More wordsmithing, and explanatory text. Work on the IANA and the security considerations sections.

A.5. draft-liman-tld-names-02

Wordsmithing and rearrangement of text following discussions with Joe Abley, Tina Dam, Thomas Narten and Andrew Sullivan. Incorporated revised ABNF and associated specification from Patrik Faltstrom. Tightened definitions and introduced the term "DNS-Label" to avoid ambiguity with various other uses of the word "label".

A.6. draft-liman-tld-names-01

Substantial comments and improvements supplied by Thomas Narten and John Klensin. Decided to go for a minimal change approach. Also noted that U-labels have to be letters due to jumping digit problem. Rewritten major parts.

A.7. draft-liman-tld-names-00

First cut. Prompted by Olafur Gudmundsson and Tina Dam.
A.8. Version Identification Tag

$Id: draft-liman-tld-names.xml,v 1.40 2011/04/12 08:20:42 liman Exp $
Authors’ Addresses

Lars-Johan Liman
Netnod Internet Exchange
Box 30194
SE-104 25 Stockholm
Sweden

Email: liman@netnod.se
URI: http://www.netnod.se/

Joe Abley
ICANN
4676 Admiralty Way
Suite 330
Marina del Rey 90292
USA

Email: joe.abley@icann.org
URI: http://www.icann.org/