Internet Draft
draft-ietf-idn-idna-01.txt

Expires in six months

Patrik Faltstrom
Cisco
Paul Hoffman
IMC & VPNC

# Internationalizing Host Names In Applications (IDNA)

## Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

## Abstract

The current DNS infrastructure does not provide a way to use internationalized host names (IDN). This document describes a mechanism that requires no changes to any DNS server or resolver that will allow internationalized host names to be used by end users with changes only to applications. It allows flexibility for user input and display, and assures that host names that have non-ASCII characters are not sent to DNS servers or resolvers.

## 1. Introduction

In the discussion of IDN solutions, a great deal of discussion has focused on transition issues and how IDN will work in a world where not all of the components have been updated. Earlier proposed solutions require that user applications, resolvers, and DNS servers to be updated in order for a user to use an internationalized host name. Instead of this requirement for widespread updating of all components, the current proposal is that only user applications be updated; no changes are needed to the DNS protocol or any DNS servers or the resolvers on user's computers.

This document is being discussed on the ietf-idna@mail.apps.ietf.org mailing list. To subscribe, send a message to ietf-idna-request@mail.apps.ietf.org with the single word "subscribe" in the body of the message.

### 1.1 Design philosophy

To date, the proposals for IDN protocols have required that DNS servers be updated to handle internationalized host names. Because of this, the person who wanted to use an internationalized host name had to be sure that their request went to a DNS server that was updated for IDN. Further, that server could only send queries to other servers that had been updated for IDN because the queries contain new protocol elements to differentiate IDN name parts from current host parts. In addition, these proposals require that resolvers must be updated to use the new protocols, and in most cases the applications would need to be updated as well.

Updating all (or even a significant percentage) of the DNS servers in the world will be difficult, to say the least. Because of this, we have designed a protocol that requires no updating of any name servers. IDNA still requires the updating of applications, but once a user has updated these, she or he could immediately start using internationalized host names. The cost of implementing IDN would thus be much lower, and the speed of implementation will be much higher.

### 1.2 Terminology

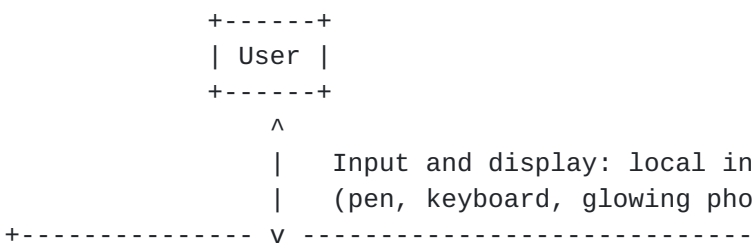
The key words "MUST", "SHALL", "REQUIRED", "SHOULD", "RECOMMENDED", and "MAY" in this document are to be interpreted as described in RFC 2119 [RFC2119].

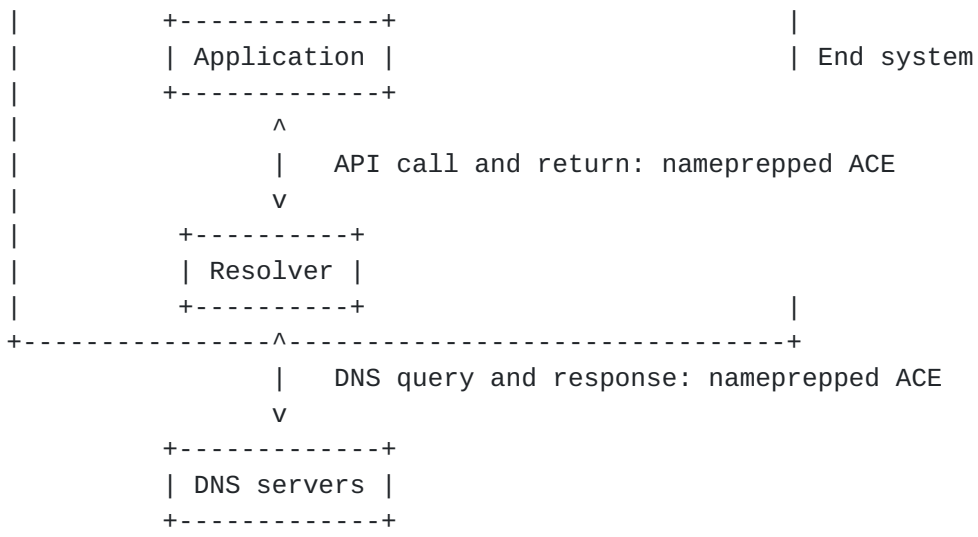
## 2. Structural Overview

In IDNA, users' applications are updated to perform the processing needed to input internationalized host names from users, display internationalized host names that are returned from the DNS to users, and process the inputs and outputs from the DNS.

### 2.1 Interfaces between DNS components in IDNA

The interfaces in IDNA can be represented pictorially as:






This document uses the generic term "ACE" for an ASCII-compatible encoding. After the IDN Working Group has chosen a specific ACE, this document will be updated to refer to just that single ACE. Until that time, an implementor creating experimental software must choose an ACE to use, such as RACE or LACE or DUDE.

### 2.1.1 Users and applications

Applications can accept host names using any character set or sets desired by the application developer, and can display host names in any charset. That is, this protocol does not affect the interface between users and applications.

An IDNA-aware application can accept and display internationalized host names in two formats: the internationalized character set(s) supported by the application, and in an ACE. Applications MAY allow ACE input and output, but are not encouraged to do so except as an interface for advanced users, possibly for debugging. ACE encoding is opaque and ugly, and should thus only be exposed to users who absolutely need it. The optional use, especially during a transition period, of ACE encodings in the user interface is described in section 3. Since ACE can be rendered either as the encoded ASCII glyphs or the proper decoded character glyphs, the rendering engine for an application SHOULD have an option for the user to select the preferred display; if it does, rendering the ACE SHOULD NOT be the default.

### 2.1.2 Applications and resolvers

Applications communicate with resolver libraries through a programming interface (API). Typically, the IETF does not standardize APIs, although there are non-standard APIs specified for IPv6. This protocol does not specify a specific API, but instead specifies only the input and output formats of the host names to the resolver library.

Before converting the name parts into ACE, the application MUST prepare each name part as specified in [NAMEPREP]. The application MUST use ACE for the name parts that are sent to the resolver, and will always get name parts encoded in ACE from the resolver.

IDNA-aware applications MUST be able to work with both non-internationalized host name parts (those that conform to [STD13] and [STD3]) and internationalized host name parts. An IDNA-aware application that is resolving a non-internationalized host name parts MUST NOT do any preparation or conversion to ACE on any non-internationalized name part.

### 2.1.3 Resolvers and DNS servers

An operating system might have a set of libraries for converting host names to nameprepped ACE. The input to such a library might be in one or more charsets that are used in applications (UTF-8 and UTF-16 are likely candidates for almost any operating system, and script-specific charsets are likely for localized operating systems). The output would be either the unchanged name part (if the input already conforms to [STD13] and [STD3]), or the nameprepped, ACE-encoded name part.

DNS servers MUST use the ACE format for internationalized host name parts.

If a signalling system which makes negotiation possible between old and new DNS clients and servers is standardized in the future, the encoding of the query in the DNS protocol itself can be changed from ACE to something else, such as UTF-8. The question whether or not this should be used is, however, a separate problem and is not discussed in this memo.

### 2.1.4 Avoiding exposing users to the raw ACE encoding

All applications that might show the user a host name that was received from a gethostbyaddr or other such lookup SHOULD update as soon as possible in order to prevent users from seeing the ACE. However, this is not considered a big problem because so few applications show this type of resolution to users.

## 3. Name Server Considerations

It is imperative that there be only one encoding for a particular host name. ACE is an encoding for host name parts that use characters outside those allowed for host names [STD13]. Thus, a primary master name server MUST NOT contain an ACE-encoded name that decodes to a host name that is allowed in [STD13] and [STD3].

Name servers MUST NOT have any records with host names that contain internationalized name parts unless those name parts have be prepared according to [NAMEPREP]. If names that are not legal in [NAMEPREP] are

passed to an application, it will result in an error being passed to the application with no error being reported to the name server. Further, no application will ever ask for a name that is not legal in [NAMEPREP] because requests always go through [NAMEPREP] before getting to the DNS.

The host name data in zone files (as specified by section 5 of RFC 1035) MUST be both nameprepped and ACE encoded.

## 4. Root Server Considerations

Because there are no changes to the DNS protocols, adopting this protocol has no effect on the root servers.

## 5. Security Considerations

Much of the security of the Internet relies on the DNS. Thus, any change to the characteristics of the DNS can change the security of much of the Internet.

Host names are used by users to connect to Internet servers. The security of the Internet would be compromised if a user entering a single internationalized name could be connected to different servers based on different interpretations of the internationalized host name.

Because this document normatively refers to [NAMEPREP], it includes the security considerations from that document as well.

## 6. References

[NAMEPREP] Paul Hoffman & Marc Blanchet, "Preparation of Internationalized Host Names", draft-ietf-idn-nameprep.

[RFC2119] Scott Bradner, "Key words for use in RFCs to Indicate Requirement Levels", March 1997, RFC 2119.

[RFC2279] Francois Yergeau, "UTF-8, a transformation format of ISO 10646", January 1998, RFC 2279.

[STD3] Bob Braden, "Requirements for Internet Hosts -- Communication Layers" (RFC 1122) and "Requirements for Internet Hosts -- Application and Support" (RFC 1123), STD 3, October 1989.

[STD13] Paul Mockapetris, "Domain names - concepts and facilities" (RFC 1034) and "Domain names - implementation and specification" (RFC 1035, STD 13, November 1987.

## B. Changes from the -00 draft

Throughout: Changed "RACE" to "ACE" and removed RACE-specific wording.

1: Added pointer to the mailing list.

1.3: Removed the [IDNCOMP] comparison section.

2.1: Added the last paragraph discussing ACEs.

2.1.2: Changed the discussion of IPv6 APIs to say they are not standards. Added reference to [STD3].

2.1.3: Added reference to [STD3]. Removed the sentence about making things smaller.

3: Added reference to [STD3].

6: Added [STD3], and updated the reference info for [STD13]. Removed [IDNCOMP].

## C. Authors' Addresses

Patrik Faltstrom
Cisco Systems
Arstaangsvagen 31 J
S-117 43 Stockholm
Sweden
paf@cisco.com

Paul Hoffman
Internet Mail Consortium and VPN Consortium
127 Segre Place
Santa Cruz, CA  95060  USA
phoffman@imc.org

--
Patrik F ltstr÷m <paf@cisco.com>                    Cisco Systems
Consulting Engineer                                Office of the CSO
Phone: (Stockholm) +46-8-4494212          (San Jose) +1-408-525-8509
        PGP: 2DFC AAF6 16F0 F276 7843  2DC1 BC79 51D9 7D25 B8DC