                    String Profile for iSCSI Names

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.

Copyright Notice

Abstract

   The iSCSI protocol provides a way for hosts to access SCSI devices
   over an IP network. The iSCSI end-points, called initiators and
   targets, each have a globally-unique name that must be transcribable,
   as well as easily compared.

   This document describes how to prepare internationalized iSCSI names
   to increase the likelihood that name input and comparison work in
   ways that make sense for typical users throughout the world.

1. **Introduction**

   The iSCSI protocol [iSCSI] provides a way for hosts to access SCSI
   [SAM2] devices over an IP network. The iSCSI end-points, called
   initiators and targets, each have a globally-unique name, defined in
   [NDT].

   An iSCSI name is a string of UTF-8 [RFC2044] characters that includes
   a type designator, a naming authority based on domain names, and a
   unique part within the naming authority. The unique part may be
   generated based on anything the naming authority deems useful, and
   may include user input.

   These names may need to be transcribed (sent between two
   administrators via email, voice, paper, etc), so a case- insensitive
   comparison would be desirable. However, these names must often be
   compared by initiator and target implementations, most of which are
   done in simple, embedded software. This makes case-sensitive
   comparison highly desirable for these implementors.

   However, a completely case-sensitive implementation would result in
   identifiers such as "example-name" and "Example-Name" being
   different, which could lead to confusion as these names are
   transcribed.

   The goal, then, is to generate iSCSI names that can be transcribed
   and entered by users, and also compared byte-for-byte, with minimal
   confusion. To attain these goals, iSCSI names are generalized using
   a normalized character set (converted to lower case or equivalent),
   with no white space allowed, and very limited punctuation.

For those using only ASCII characters (U+0000 to U+007F), the
following characters are allowed:

- ASCII dash character ('-' = U+002d)
- ASCII dot character ('.' = U+002e)
- ASCII colon character (':' = U+003a)
- ASCII lower-case characters ('a'..'z' = U+0061..U+007a)
- ASCII digit characters ('0'..'9' = U+0030..U+0039)

In addition, any upper-case characters input via a user interface
should be mapped to their lower-case equivalents.

This document specifies the valid character set for iSCSI names,
along with the rules for normalizing and generating iSCSI names based
on user input or other information that contains international
characters.

In particular, it defines the following, as required by [RFC3454]:

- The intended applicability of the profile: internationalized iSCSI
  names.

- The character repertoire that is the input and output to
  stringprep: Unicode 3.2, specified in section 3.

- The mappings used: specified in section 4.

- The Unicode normalization used: specified in section 5.

- The characters that are prohibited as output: specified in section
  6.

This profile MUST be used with the iSCSI protocol.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119].

Examples in this document use the notation for code points and names
from the Unicode Standard [Unicode3.2] and ISO/IEC 10646 [ISO10646].
For example, the letter "a" may be represented as either "U+0061" or
"U+0061" or
"LATIN SMALL LETTER A". In the lists of prohibited characters, the
"U+" is left off to make the lists easier to read. The comments for
character ranges are shown in square brackets (such as "[SYMBOLS]")
and do not come from the standards.

3. Character Repertoire

This profile uses Unicode 3.2, as defined in [RFC3454] Appendix A.

4. Mapping

This profile specifies mapping using the following tables from
[RFC3454]. The following mapping tables MUST be used when generating
iSCSI names from Unicode characters.

    Table B.1
    Table B.2

5. Normalization

Unicode normalization form KC MUST be used with this profile, as
described in [RFC3454].

6. Prohibited Output

This profile specifies prohibiting using the following tables from
[stringprep]. Characters appearing within these tables MUST NOT be
used within an iSCSI name.

    Table C.1.1
    Table C.1.2
    Table C.2.1
    Table C.2.2
    Table C.3
    Table C.4
    Table C.5
    Table C.6
    Table C.7
    Table C.8
    Table C.9

Important note: this profile MUST be used with the iSCSI protocol.
The iSCSI protocol has additional naming rules that are checked
outside of this profile.

In addition, this profile adds the following prohibitions. The full
set of prohibited characters are those from the tables above plus
those listed individually below.

6.1. Inappropriate Characters from Common Input Mechanisms

   u+3002 is used as if it were u+002e in many domain name input
   mechanisms used by applications, particularly in Asia. The character
   u+3002 MUST NOT be used in an iSCSI name.

      3002; ideographic full stop

6.2. Currently-prohibited ASCII characters

   Some of the ASCII characters that are currently prohibited in iSCSI
   names by [NDT] are also used in protocol elements such as URIs. Some
   examples are described in [RFC2396] and [RFC2732]. Note that there
   are many other RFCs that define additional URI schemes.

   The other characters in the range U+0000 to U+007F that are not
   currently allowed are prohibited in iSCSI names to reserve them for
   future use in protocol elements. Note that the dash (U+002D), dot
   (U+002E), and colon (U+003A) are not prohibited.

   The following characters MUST NOT be used in iSCSI names:

      0000-002C; [ASCII CONTROL CHARACTERS and SPACE through ,]
      002F; [ASCII /]
      003B-0040; [ASCII ; through @]
      005B-0060; [ASCII [through `]
      007B-007F; [ASCII { through DEL]

7. Bidirectional Characters

   This profile specifies checking bidirectional strings as described in
   [RFC3454] section 6.

8. Unassigned Code Points in Internationalized Domain Names

   If the processing in [iSCSI] specifies that a list of unassigned code
   points be used, the system uses table A.1 from [RFC3454] as its list
   of unassigned code points.

9. Security Considerations

   ISO/IEC 10646 has many characters that look similar. In many cases,
   users of security protocols might do visual matching, such as when
   comparing the names of trusted third parties. This profile does

   nothing to map similar-looking characters together.

   iSCSI names may be used by an initiator to verify that a target it
   has discovered is the correct one, and by a target to verify that an
   initiator is to be allowed access. If these names are interpreted
   and compared differently by different iSCSI implementations, an
   initiator could gain access to the wrong target, or could be denied
   access to a legitimate target.

10. IANA Considerations

   This is a profile of stringprep. When it has been reviewed and
   approved by the IESG, it should be registered in the IANA "Stringprep
   Profiles" registry. This process is described in the IANA
   Considerations section of [RFC3454].

11. Summary

   This document describes a stringprep profile to be used with programs
   generating names for iSCSI initiators and targets.

12. Normative References

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

[RFC3454]   Paul Hoffman and Marc Blanchet, "Preparation of
            Internationalized Strings ("stringprep")", RFC 3454,
            December 2002.

[iSCSI]   J. Satran, et. al. "iSCSI", Work In Progress, draft-ietf-
            ips-iscsi-20.txt, January 2003.

13. Informative References

[NDT]   K. Voruganti, et. al. "iSCSI Naming and Discovery", Work in
            Progress, draft-ietf-ips-iscsi-name-disc-09, March 2003.

[RFC3491]   P. Hoffman and M. Blanchet, "Nameprep: A Stringprep Profile
            for Internationalized Domain Names", RFC 3491, March 2003.

[SAM2]   ANSI T10. "SCSI Architectural Model 2", March 2000.

[RFC2044]   F. Yergeau, "UTF-8, a transformation format of Unicode and
             ISO 10646", RFC 2044, October 1996.

[RFC2396]   T. Berners-Lee, R. Fielding, L. Masinter, "Uniform Resource
             Identifiers", RFC 2396, August 1998.

[RFC2732]   R. Hindon, B. Carpenter, L Masinter, "Format for Literal
             IPv6 Addresses in URLs", RFC 2732, December 1999.

[Unicode3.2]
             The Unicode Standard, Version 3.2.0: The Unicode Consortium.
             The Unicode Standard, Version 3.2.0 is defined by The
             Unicode Standard, Version 3.0 (Reading, MA, Addison-Wesley,
             2000. ISBN 0-201-61633-5), as amended by the Unicode
             Standard Annex #27: Unicode 3.1
             (http://www.unicode.org/unicode/reports/tr27/) and by the
             Unicode Standard Annex #28: Unicode 3.2
             (http://www.unicode.org/unicode/reports/tr28/).

[ISO10646]   ISO/IEC 10646-1:2000. International Standard -- Information
             technology -- Universal Multiple-Octet Coded Character Set
             (UCS) -- Part 1: Architecture and Basic Multilingual Plane.

14. Author Contact Information

        Mark Bakke
        Cisco Systems, Inc.
        6450 Wedgwood Road
        Maple Grove, MN
        USA 55311

        Voice: +1 763-398-1000
        E-Mail: mbakke@cisco.com

15. Full Copyright Notice

Acknowledgement