Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 21, 2011

M. Blanchet
Viagenie
A. Sullivan

October 18, 2010

Stringprep Revision Problem Statement
draft-ietf-precis-problem-statement-00.txt

Abstract

Using Unicode codepoints in protocol strings that expect comparison with other strings [[anchor1: The WG will need to decide whether "other strings" is too broad. In particular, what about protocol slots that can take strings other than plain ASCII? --ajs@shinkuro.com]] requires preparation of the string that contains the Unicode codepoints. Internationalizing Domain Names in Applications (IDNA2003) defined and used Stringprep and Nameprep. Other protocols subsequently defined Stringprep profiles. A new approach different from Stringprep and Nameprep is used for a revision of IDNA2003 (called IDNA2008). Other Stringprep profiles need to be similarly updated or a replacement of Stringprep need to be designed. This document outlines the issues to be faced by those designing a Stringprep replacement.

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 April 21, 2011.

Table of Contents

1. Introduction

Internationalizing Domain Names in Applications (IDNA2003) [RFC3490], [RFC3491], [RFC3492], [RFC3454] described a mechanism for encoding UTF-8 labels making up Internationalized Domain Names (IDNs) as standard DNS labels. The labels were processed using a method called Nameprep [RFC3491] and Punycode [RFC3492]. That method was specific to IDNA2003, but is generalized as Stringprep [RFC3454]. The general mechanism can be used to help other protocols with similar needs, but with different constraints than IDNA2003.

Stringprep defines a framework within which protocols define their Stringprep profiles. Known IETF specifications using Stringprep are listed below:

- The Nameprep profile [RFC3490] for use in Internationalized Domain Names (IDNs);
- NFSv4 [RFC3530] and NFSv4.1 [RFC5661];
- The iSCSI profile [RFC3722] for use in Internet Small Computer Systems Interface (iSCSI) Names;
- EAP [RFC3748];
- The Nodeprep and Resourceprep profiles [RFC3920] for use in the Extensible Messaging and Presence Protocol (XMPP), and the XMPP to CPIM mapping [RFC3922];
- The Policy MIB profile [RFC4011] for use in the Simple Network Management Protocol (SNMP);
- The SASLprep profile [RFC4013] for use in the Simple Authentication and Security Layer (SASL), and SASL itself [RFC4422];
- TLS [RFC4279];
- IMAP4 using SASLprep [RFC4314];
- The trace profile [RFC4505] for use with the SASL ANONYMOUS mechanism;
- The LDAP profile [RFC4518] for use with LDAP [RFC4511] and its authentication methods [RFC4513];
- Plain SASL using SASLprep [RFC4616];
- NNTP using SASLprep [RFC4643];
- PKIX subject identification using LDAPprep [RFC4683];
- Internet Application Protocol Collation Registry [RFC4790];
- SMTP Auth using SASLprep [RFC4954];
- POP3 Auth using SASLprep [RFC5034];
- TLS SRP using SASLprep [RFC5054];
- IRI and URI in XMPP [RFC5122];
- PKIX CRL using LDAPprep [RFC5280];
- IAX using Nameprep [RFC5456];
- SASL SCRAM using SASLprep [RFC5802];
- Remote management of Sieve using SASLprep [RFC5804];

- The i;unicode-casemap Unicode Collation [RFC5051].

There turned out to be some difficulties with IDNA2003, documented in [RFC4690]. These difficulties led to a new IDN specification, called IDNA2008 [RFC5890], [RFC5891], [RFC5892], [RFC5893]. Additional background and explanations of the decisions embodied in IDNA2008 is presented in [RFC5894]. One of the effects of IDNA2008 is that Nameprep and Stringprep are not used at all. Instead, an algorithm based on Unicode properties of codepoints is defined. That algorithm generates a stable and complete table of the supported Unicode codepoints. This algorithm is based on an inclusion-based approach, instead of the exclusion-based approach of Stringprep/Nameprep.

This document lists the shortcomings and issues found by protocols listed above that defined Stringprep profiles. It also lists some early conclusions and requirements for a potential replacement of Stringprep.

2. Usage and Issues of Stringprep

2.1. Issues raised during newprep BOF

During IETF 77, a BOF discussed the current state of the protocols that have defined Stringprep profiles [NEWPREP]. The main conclusions are :

- Stringprep is bound to a specific version of Unicode: 3.2. Stringprep has not been updated to new versions of Unicode. Therefore, the protocols using Stringprep are stuck to Unicode 3.2.
- The protocols need to be updated to support new versions of Unicode. The protocols would like to not be bound to a specific version of Unicode, but rather have better Unicode agility in the way of IDNA2008. This is important partly because it is usually impossible for an application to require Unicode 3.2; the application gets whatever version of Unicode is available on the host.
- The protocols require better bidirectional support (bidi) than currently offered by Stringprep.
- If the protocols are updated to use a new version of Stringprep or another framework, then backward compatibility is an important requirement. For example, Stringprep is based on and may use NFKC [UAX15], while IDNA2008 mostly uses NFC [UAX15].
- Protocols use each other; for example, a protocol can use user identifiers that are later passed to SASL, LDAP or another authentication mechanism. Therefore, common set of rules or classes of strings are preferred over specific rules for each protocol.

Protocols that use Stringprep profiles use strings for different purposes:

- XMPP uses a different Stringprep profile for each part of the XMPP address (JID): a localpart which is similar to a username and used for authentication, a domainpart which is a domain name and a resource part which is less restrictive than the localpart.
- iSCSI uses a Stringprep profile for the IQN, which is very similar to (often is) a DNS domain name.
- SASL and LDAP uses a Stringprep profile for usernames.
- LDAP uses a set of Stringprep profiles.

During the newprep BOF, it was the consensus of the attendees that it would be highly preferable to have a replacement of Stringprep, with similar characteristics to IDNA2008. That replacement should be defined so that the protocols could use internationalized strings without a lot of specialized internationalization work, since internationalization expertise is not available in the respective protocols or working groups.

2.2. Specific issues with particular Stringprep profiles

[[anchor6: This section is where issues raised in the individual profile reviews goes. A review of the WG trac state on 2010-10-06 of the tracker suggests those reviews haven't happened yet. --ajs@shinkuro.com]]

2.3. Inclusion vs. exclusion of characters

One of the primary changes of IDNA2008 is in the way it approaches Unicode characters. IDNA2003 created an explicit list of excluded or mapped-away characters; anything in Unicode 3.2 that was not so listed could be assumed to be allowed under the protocol. IDNA2008 begins instead from the assumption that characters are disallowed, and then relies on Unicode properties to derive whether a given character actually is allowed in the protocol.

Moreover, there is more than one class of "allowed in the protocol". While some characters are simply disallowed, some are allowed only in certain contexts. The reasons for the context-dependent rules have to do with the way some characters are used. For instance, the ZERO WIDTH JOINER and ZERO WIDTH NON-JOINER characters (ZWJ, U+200D and ZWNJ, U+200C) are allowed with contextual rules because they are required in some circumstances, yet are considered punctuation by Unicode and would therefore be DISALLOWED under the usual IDNA2008 derivation rules.

The working group needs to decide whether similar contextual cases need to be supported.

2.4. Stringprep and NFKC

Stringprep profiles may use normalization. If they do, they use NFKC [UAX15]. It is not clear that NFKC is the right normalization to use in all cases. In [UAX15], there is the following observation regarding Normalization Forms KC and KD: "It is best to think of these Normalization Forms as being like uppercase or lowercase mappings: useful in certain contexts for identifying core meanings, but also performing modifications to the text that may not always be appropriate." For things like the spelling of users' names, then, NKFC may not be the best form to use. At the same time, one of the nice things about NFKC is that it deals with the width of characters that are otherwise similar, by canonicalizing half-width to full-width. This mapping step can be crucial in practice. The WG will need to analyze the different use profiles and consider whether NFKC or NFC is a better normalization for each profile.

2.5. Case mapping

In IDNA2003, labels are always mapped to lower case before the Punycode transformation. In IDNA2003, there is no mapping at all: input is either a valid U-label or it is not. At the same time, upper-case characters are by definition not valid U-labels, because they fall into the Unstable category (category B) of [RFC5892].

If there are protocols that require upper and lower cases be preserved, then the analogy with IDNA2008 will break down. The working group will need to decide whether there are any cases that require upper case, and what to do about it if so.

2.6. Whether to use ASCII-compatible encoding

The development of IDNA2008 depended on the notion that there was a narrow repertoire of reasonable traditional labels, and what was necessary was to internationalize that repertoire rather than to incorporate any characters into domain name labels. More exactly, the idea was to internationalize the traditional hostname rules (the "LDH rule". See [RFC4690], section 5.1.). Efforts to internationalize email ([RFC5336]) have started from different assumptions. The email example suggests that in some cases, the right answer might be to internationalize the target protocol rather than to depend on a technology to ensure protocol slots can use only ASCII. The working group will need to determine which approach is correct for the different use-cases.

2.7. Issues with delimiters

There are two kinds of issues to address with delimiters. First, exactly where a delimiter will appear on the screen when dealing with bidirectional parts of a string can be extremely surprising. In the case of IDNA2008, just what to do in these cases remains a display issue (there is no question about the wire format, because the wire format is an A-label and it is always left to right).

Second, there is the question of whether to include different kinds of protocol separators. For instance, FULL STOP, U+002E (.) may not be available on all keyboards. In addition, in some languages there is more than one full stop which are variants of one another. The working group will need to decide how to handle such cases: whether there will be a mapping, some restrictions, or something else.

3. Considerations for Stringprep replacement

The above suggests the following direction for the working group:

- A stringprep replacement should be defined.
- The replacement should take an approach similar to IDNA2008, in that it enables Unicode agility.
- Protocols share similar characteristics of strings. Therefore, defining i18n preparation algorithms for a (small) set of string classes may be sufficient for most cases and provides the coherence among a set of protocol friends.
- The sets of string classes need to be evaluated for the following properties:
  - the normalization needed (NFC vs NFKC);
  - whether case-folding, case preservation, and case-insensitive matching is needed;
  - what restrictions on input are reasonable for the class (i.e. whether there is something like an "LDH rule" for the class), or whether the ASCII-only input in the protocol slot is lightly constrained;
  - the extent to which bidi considerations are important for the class.

Existing deployments already depend on Stringprep profiles. Therefore, the working group will need to consider the effects of any new strategy on existing deployments. By way of comparison, it is worth noting that some characters were acceptable in IDNA labels under IDNA2003, but are not protocol-valid under IDNA2008 (and conversely). Different implementers may make different decisions about what to do in such cases; this could have interoperability effects. The working group will need to trade better support for different linguistic environments against the potential side effects

of backward incompatibility.

4. Security Considerations

This document merely states what problems are to be solved, and does not define a protocol. There are undoubtedly security implications of the particular results that will come from the work to be completed.

5. IANA Considerations

This document has no actions for IANA.

6. Discussion home for this draft

This document is intended to define the problem space discussed on the precis@ietf.org mailing list.

7. Informative References

[NEWPREP] "Newprep BoF Meeting Minutes", March 2010.

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

[RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.

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

[RFC3492] Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.

[RFC3530] Shepler, S., Callaghan, B., Robinson, D., Thurlow, R., Beame, C., Eisler, M., and D. Noveck, "Network File System (NFS) version 4 Protocol", RFC 3530, April 2003.

[RFC3722] Bakke, M., "String Profile for Internet Small Computer Systems Interface (iSCSI) Names", RFC 3722, April 2004.

[RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, "Extensible Authentication Protocol (EAP)", RFC 3748, June 2004.

[RFC3920] Saint-Andre, P., Ed., "Extensible Messaging and Presence Protocol (XMPP): Core", RFC 3920, October 2004.

[RFC3922] Saint-Andre, P., "Mapping the Extensible Messaging and Presence Protocol (XMPP) to Common Presence and Instant Messaging (CPIM)", RFC 3922, October 2004.

[RFC4011] Waldbusser, S., Saperia, J., and T. Hongal, "Policy Based Management MIB", RFC 4011, March 2005.

[RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", RFC 4013, February 2005.

[RFC4279] Eronen, P. and H. Tschofenig, "Pre-Shared Key Ciphersuites for Transport Layer Security (TLS)", RFC 4279, December 2005.

[RFC4314] Melnikov, A., "IMAP4 Access Control List (ACL) Extension", RFC 4314, December 2005.

[RFC4422] Melnikov, A. and K. Zeilenga, "Simple Authentication and Security Layer (SASL)", RFC 4422, June 2006.

[RFC4505] Zeilenga, K., "Anonymous Simple Authentication and Security Layer (SASL) Mechanism", RFC 4505, June 2006.

[RFC4511] Sermersheim, J., "Lightweight Directory Access Protocol (LDAP): The Protocol", RFC 4511, June 2006.

[RFC4513] Harrison, R., "Lightweight Directory Access Protocol (LDAP): Authentication Methods and Security Mechanisms", RFC 4513, June 2006.

[RFC4518] Zeilenga, K., "Lightweight Directory Access Protocol (LDAP): Internationalized String Preparation", RFC 4518, June 2006.

[RFC4616] Zeilenga, K., "The PLAIN Simple Authentication and Security Layer (SASL) Mechanism", RFC 4616, August 2006.

[RFC4643] Vinocur, J. and K. Murchison, "Network News Transfer Protocol (NNTP) Extension for Authentication", RFC 4643, October 2006.

[RFC4683] Park, J., Lee, J., Lee, H., Park, S., and T. Polk, "Internet X.509 Public Key Infrastructure Subject Identification Method (SIM)", RFC 4683, October 2006.

[RFC4690] Klensin, J., Faltstrom, P., Karp, C., and IAB, "Review and Recommendations for Internationalized Domain Names (IDNs)", RFC 4690, September 2006.

[RFC4790] Newman, C., Duerst, M., and A. Gulbrandsen, "Internet Application Protocol Collation Registry", RFC 4790, March 2007.

[RFC4954] Siemborski, R. and A. Melnikov, "SMTP Service Extension for Authentication", RFC 4954, July 2007.

[RFC5034] Siemborski, R. and A. Menon-Sen, "The Post Office Protocol (POP3) Simple Authentication and Security Layer (SASL) Authentication Mechanism", RFC 5034, July 2007.

[RFC5051] Crispin, M., "i;unicode-casemap - Simple Unicode Collation Algorithm", RFC 5051, October 2007.

[RFC5054] Taylor, D., Wu, T., Mavrogiannopoulos, N., and T. Perrin, "Using the Secure Remote Password (SRP) Protocol for TLS Authentication", RFC 5054, November 2007.

[RFC5122] Saint-Andre, P., "Internationalized Resource Identifiers (IRIs) and Uniform Resource Identifiers (URIs) for the Extensible Messaging and Presence Protocol (XMPP)", RFC 5122, February 2008.

[RFC5280] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., and W. Polk, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", RFC 5280, May 2008.

[RFC5336] Yao, J. and W. Mao, "SMTP Extension for Internationalized Email Addresses", RFC 5336, September 2008.

[RFC5456] Spencer, M., Capouch, B., Guy, E., Miller, F., and K. Shumard, "IAX: Inter-Asterisk eXchange Version 2", RFC 5456, February 2010.

[RFC5661] Shepler, S., Eisler, M., and D. Noveck, "Network File System (NFS) Version 4 Minor Version 1 Protocol", RFC 5661, January 2010.

[RFC5802] Newman, C., Menon-Sen, A., Melnikov, A., and N. Williams,

"Salted Challenge Response Authentication Mechanism (SCRAM) SASL and GSS-API Mechanisms", RFC 5802, July 2010.

[RFC5804] Melnikov, A. and T. Martin, "A Protocol for Remotely Managing Sieve Scripts", RFC 5804, July 2010.

[RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, August 2010.

[RFC5891] Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", RFC 5891, August 2010.

[RFC5892] Faltstrom, P., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892, August 2010.

[RFC5893] Alvestrand, H. and C. Karp, "Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA)", RFC 5893, August 2010.

[RFC5894] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale", RFC 5894, August 2010.

[UAX15] "Unicode Standard Annex #15: Unicode Normalization Forms", UAX 15, September 2009.

Authors' Addresses

Marc Blanchet
Viagenie
2600 boul. Laurier, suite 625
Quebec, QC G1V 4W1
Canada

Email: Marc.Blanchet@viagenie.ca
URI: http://viagenie.ca

Andrew Sullivan
519 Maitland St.
London, ON N6B 2Z5
Canada

Email: ajs@crankycanuck.ca