       **Preparation and Comparison of Internationalized Strings Representing
                    Simple User Names and Passwords
                    draft-melnikov-precis-saslprepbis-02**

Abstract

   This document describes how to handle Unicode strings representing
   simple user names and passwords, primarily for purposes of
   comparison. This profile is intended to be used by Simple
   Authentication and Security Layer (SASL) mechanisms (such as PLAIN
   and SCRAM-SHA-1), as well as other protocols that exchange simple
   user names or user passwords. This document obsoletes RFC 4013.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

1.1. Overview

   User names and passwords are used pervasively in authentication and
   authorization on the Internet. To increase the likelihood that the
   input and comparison of user names and passwords will work in ways
   that make sense for typical users throughout the world, this document
   defines rules for preparing and comparing internationalized strings
   that represent simple user names and passwords.

   The algorithms defined in this document assume that all strings are
   comprised of characters from the Unicode character set [UNICODE].

   The algorithms are designed for use in Simple Authentication and
   Security Layer (SASL) [RFC4422] mechanisms, such as PLAIN [RFC4616]
   and SCRAM-SHA-1 [RFC5802]. However, they might be applicable
   wherever simple user names or passwords are used. This profile is
   not intended for use in preparing strings that are not simple user
   names (e.g., email addresses, DNS domain names, LDAP distinguished
   names), nor in cases where identifiers or secrets are not character
   data or require different handling (e.g., case folding).

   This document builds upon the PRECIS framework defined in
   [FRAMEWORK], which differs fundamentally from the stringprep
   technology [RFC3454] used in SASLprep [RFC4013]. The primary
   difference is that stringprep profiles allowed all characters except
   those which were explicitly disallowed, whereas PRECIS profiles
   disallow all characters except those which are explicitly allowed
   (this "inclusion model" was originally used for internationalized
   domain names in [RFC5891]; see [RFC5894] for further discussion). It
   is important to keep this distinction in mind when comparing the
   technology defined in this document to SASLprep [RFC4013].

   This document obsoletes RFC 4013.

1.2. Terminology

   Many important terms used in this document are defined in
   [FRAMEWORK], [RFC4422], [RFC5890], [RFC6365], and [UNICODE]. The
   term "non-ASCII" space refers to any Unicode code point with a
   general category of "Zs", with the exception of U+0020 (here called
   "ASCII space").

   As used here, the term "password" is not literally limited to a word;
   i.e., a password could be a passphrase consisting of more than one
   word, perhaps separated by spaces or other such characters.

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

2. Simple User Names

2.1. Definition

Some SASL mechanisms (e.g., CRAM-MD5, DIGEST-MD5, and SCRAM) specify
that the authentication identity used in the context of such
mechanisms is a "simple user name" (see Section 2 of [RFC4422] as
well as [RFC4013]). However, the exact form of a simple user name in
any particular mechanism or deployment thereof is a local matter, and
a simple user name does not necessarily map to an application
identifier such as the localpart of an email address.

For purposes of preparation and comparison of authentication
identities, this document specifies that a simple user name is a
string of [UNICODE] code points, encoded using UTF-8 [RFC3629], and
structured as an ordered sequence of "simpleparts" (where the
complete simple user name can consist of a single simplepart or a
space-separated sequence of simpleparts).

Therefore the syntax for a simple user name is defined as follows
using the Augmented Backus-Naur Form (ABNF) as specified in
[RFC5234].

```
simpleusername = simplepart [1*(1*SP simplepart)]
simplepart     = 1*(namepoint)
                   ;
                   ; a "namepoint" is a UTF-8 encoded
                   ; Unicode code point that conforms to
                   ; the "NameClass" string class defined
                   ; in draft-ietf-precis-framework
                   ;
```

2.2. Preparation

A simple user name MUST NOT be zero bytes in length. This rule is to
be enforced after any normalization or mapping of code points.

Each simplepart of a simple user name MUST be treated as follows,
where the operations specified MUST be completed in the order shown:

1. Apply Unicode Normalization Form C (NFC) to all characters.

2. Map uppercase and titlecase characters to their lowercase
    equivalents.

3. Optionally apply additional mappings, such as those defined in
    [MAPPINGS].

4. Ensure that the resulting string conforms to the definition of
    the PRECIS NameClass.

With regard to directionality, the "Bidi Rule" provided in [RFC5893]
applies.

2.3. Migration

The rules defined in the previous section differ slightly from those
defined by the SASLprep specification [RFC4013]. Therefore,
deployments that currently use SASLprep for handling user names will
need to scrub existing data when migrating to use of the rules
defined here. In particular:

o SASLprep specified the use of Unicode Normalization Form KC
   (NFKC), whereas this usage of the PRECIS NameClass employs Unicode
   Normalization Form C (NFC). In practice this change is unlikely
   to cause significant problems, because NFKC provides methods for
   mapping Unicode code points with compatibility equivalents to
   those equivalents, whereas the PRECIS NameClass entirely disallows
   Unicode code points with compatibility equivalents. For migration
   purposes, deployments need to search their simple user names for
   Unicode code points with compatibility equivalents and map those
   code points to their compatibility equivalents.

o SASLprep mapped non-ASCII spaces to ASCII space (U+0020), whereas
   the PRECIS NameClass entirely disallows non-ASCII spaces. For
   migration purposes, deployments need to convert non-ASCII space
   characters to ASCII space in simple user names.

o SASLprep mapped the "characters commonly mapped to nothing" from
   Appendix B.1 of [RFC3454]) to nothing, whereas the PRECIS
   NameClass entirely disallows such characters, which correspond to
   the code points from the "M" category defined under Section 6.13
   of [FRAMEWORK] (with the exception of U+1806 MONGOLIAN TODO SOFT
   HYPHEN, which was commonly mapped to nothing in Unicode 3.2 but at
   the time of this writing is allowed by Unicode 6.1). For
   migration purposes, deployments need to remove code points from
   the PRECIS "M" category in simple user names.

o SASLprep allowed uppercase and titlecase characters, whereas this
   usage of the PRECIS NameClass maps uppercase and titlecase
   characters to their lowercase equivalents. For migration
   purposes, deployments can either convert uppercase and titlecase
   characters to their lowercase equivalents in simple user names
   (thus losing the case information) or preserve uppercase and
   titlecase characters and ignore the case difference when comparing
   simple user names.

Note well that all code points and blocks not explicitly allowed in
the PRECIS NameClass are disallowed; this includes private use
characters, surrogate code points, and the other code points and
blocks defined as "Prohibited Output" in Section 2.3 of RFC 4013.

3. Passwords

3.1. Definition

For purposes of preparation and comparison of passwords, this
document specifies that a password is a string of [UNICODE] code
points, encoded using UTF-8 [RFC3629], and conformant to the PRECIS
FreeClass.

Therefore the syntax for a password is defined as follows using the
Augmented Backus-Naur Form (ABNF) as specified in [RFC5234].

```
     password        = 1*(freepoint)
                        ;
                        ; a "freepoint" is a UTF-8 encoded
                        ; Unicode code point that conforms to
                        ; the "FreeClass" string class defined
                        ; in draft-ietf-precis-framework
                        ;
```

3.2. Preparation

A password MUST NOT be zero bytes in length. This rule is to be
enforced after any normalization or mapping of code points.

A password MUST be treated as follows, where the operations specified
MUST be completed in the order shown:

1. Apply Unicode Normalization Form C (NFC) to all characters.

2. Map any instances of non-ASCII space to ASCII space (U+0020).

3. Ensure that the resulting string conforms to the definition of
    the PRECIS FreeClass.

With regard to directionality, the "Bidi Rule" provided in [RFC5893]
applies.

3.3. Migration

The rules defined in the previous section differ slightly from those
defined by the SASLprep specification [RFC4013]. Depending on local
service policy, migration from RFC 4013 to this specification might
not involve any scrubbing of data (since passwords might not be
stored in the clear anyway); however, service providers need to be
aware of possible issues that might arise during migration. In
particular:

o SASLprep specified the use of Unicode Normalization Form KC
   (NFKC), whereas this usage of the PRECIS FreeClass employs Unicode
   Normalization Form C (NFC). Because NFKC is more aggressive about
   finding matches than NFC, in practice this change is unlikely to
   cause significant problems and indeed will probably result in
   fewer false positives when comparing passwords.

o SASLprep mapped the "characters commonly mapped to nothing" from
   Appendix B.1 of [RFC3454]) to nothing, whereas the PRECIS
   FreeClass entirely disallows such characters, which correspond to
   the code points from the "M" category defined under Section 6.13
   of [FRAMEWORK] (with the exception of U+1806 MONGOLIAN TODO SOFT
   HYPHEN, which was commonly mapped to nothing in Unicode 3.2 but at
   the time of this writing is allowed by Unicode 6.1).

Note well that all code points and blocks not explicitly allowed in
the PRECIS FreeClass are disallowed; this includes private use
characters, surrogate code points, and the other code points and
blocks defined as "Prohibited Output" in Section 2.3 of RFC 4013.

4. Open Issues

We need to compare the output obtained when applying the new rules
with Unicode 3.2 and Unicode 6.1 data to the output obtained when
applying the SASLprep rules with Unicode 3.2 data, then make sure
that the PRECIS Working Group and KITTEN Working Group are
comfortable with any changes to the Unicode characters that are
allowed and disallowed. (See also the migration issues described in
the foregoing sections.)

5. Security Considerations

5.1. Password/Passphrase Strength

   The ability to include a wide range of characters in passwords and
   passphrases can increase the potential for creating a strong password
   with high entropy. However, in practice, the ability to include such
   characters ought to be weighed against the possible need to reproduce
   them on various devices using various input methods.

5.2. Reuse of PRECIS

   The security considerations described in [FRAMEWORK] apply to the
   "NameClass" and "FreeClass" base string classes used in this document
   for user names and passwords, respectively.

5.3. Reuse of Unicode

   The security considerations described in [UTR39] apply to the use of
   Unicode characters in user names and passwords.

6. IANA Considerations

6.1. Use of NameClass

   The IANA shall add an entry to the PRECIS Usage Registry for reuse of
   the PRECIS NameClass in SASL, as follows:

   Application Protocol: SASL/Kerberos.
   Base Class: NameClass.
   Subclassing: No.
   Directionality: The "Bidi Rule" defined in RFC 5893 applies.
   Casemapping: Map uppercase and titlecase code points to their
      lowercase equivalents.
   Normalization: NFC.
   Specification: RFC XXXX.

6.2. Use of FreeClass

   The IANA shall add an entry to the PRECIS Usage Registry for reuse of
   the PRECIS FreeClass in SASL, as follows:

   Application Protocol: SASL/Kerberos.

      Base Class: FreeClass
      Subclassing: No.
      Directionality: The "Bidi Rule" defined in RFC 5893 applies.
      Casemapping: None.
      Normalization: NFC.
      Specification: RFC XXXX.

7. References

7.1. Normative References

   [FRAMEWORK]
               Saint-Andre, P. and M. Blanchet, "Precis Framework:
               Handling Internationalized Strings in Protocols",
               draft-ietf-precis-framework-05 (work in progress),
               August 2012.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
               Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC3629]   Yergeau, F., "UTF-8, a transformation format of ISO
               10646", STD 63, RFC 3629, November 2003.

   [RFC5234]   Crocker, D. and P. Overell, "Augmented BNF for Syntax
               Specifications: ABNF", STD 68, RFC 5234, January 2008.

   [UNICODE]   The Unicode Consortium, "The Unicode Standard, Version
               6.1", 2012,
               <http://www.unicode.org/versions/Unicode6.1.0/>.

7.2. Informative References

   [MAPPINGS]
               YONEYA, Y. and T. NEMOTO, "Mapping characters for PRECIS
               classes", draft-yoneya-precis-mappings-02 (work in
               progress), July 2012.

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

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

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

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

   [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.

   [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.

   [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.

   [RFC6365]   Hoffman, P. and J. Klensin, "Terminology Used in
               Internationalization in the IETF", BCP 166, RFC 6365,
               September 2011.

   [UTR39]   The Unicode Consortium, "Unicode Technical Report #39:
               Unicode Security Mechanisms", August 2010,
               <http://unicode.org/reports/tr39/>.

Appendix A. Differences from RFC 4013

   The following substantive modifications were made from RFC 3920.

   o A single SASLprep algorithm was replaced by two separate
      algorithms: one for user names and another for passwords.
   o The new preparation algorithms use PRECIS instead of a stringprep
      profile. The new algorithms work independenctly of Unicode
      versions.
   o As recommended in the PRECIS framwork, changed the Unicode
      normalization form from NFKC to NFC.
   o Some Unicode code points that were mapped to nothing in RFC 4013
      are simply disallowed by PRECIS.

Appendix B. **Acknowledgements**

   Thanks to Yoshiro YONEYA and Takahiro NEMOTO for their implementation
   feedback.

   This document borrows some text from RFC 4013 and RFC 6120.

Authors' Addresses

   Peter Saint-Andre
   Cisco Systems, Inc.
   1899 Wynkoop Street, Suite 600
   Denver, CO 80202
   USA

   Phone: +1-303-308-3282
   Email: psaintan@cisco.com

   Alexey Melnikov
   Isode Ltd
   5 Castle Business Village
   36 Station Road
   Hampton, Middlesex TW12 2BX
   UK

   Email: Alexey.Melnikov@isode.com