PRECIS
Internet-Draft
Obsoletes: 4013 (if approved)
Intended status: Standards Track
Expires: September 28, 2013

P. Saint-Andre
Cisco Systems, Inc.
A. Melnikov
Isode Ltd

March 27, 2013

Preparation and Comparison of Internationalized Strings Representing Simple User Names and Passwords
draft-ietf-precis-saslprepbis-01

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 passwords. This document obsoletes RFC 4013.

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 September 28, 2013.

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. (In many authentication technologies passwords are not directly compared because the actual password is used as input to an algorithm such as a hash function; however, non-ASCII code points in the input string still need to be handled correctly.)

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 strings (e.g., keys or certificates) 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 [UNICODE], encoded using UTF-8 [RFC3629], and structured either 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) or as a simplepart@domainpart (where the domainpart is an IP literal, an IPv4 address, or a fully-qualified domain name).

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 '@' domainpart
simplepart     = 1*(idpoint)
                 ;
                 ; an "idpoint" is a UTF-8 encoded
                 ; Unicode code point that conforms to
                 ; the PRECIS "IdentifierClass"
                 ;
domainpart     = IP-literal / IPv4address / ifqdn
                 ;
                 ; the "IPv4address" and "IP-literal"
                 ; rules are defined in RFC 3986, and
                 ; the first-match-wins (a.k.a. "greedy")
                 ; algorithm described in RFC 3986
                 ; applies to the matching process
                 ;
                 ; note well that reuse of the IP-literal
                 ; rule from RFC 3986 implies that IPv6
                 ; addresses are enclosed in square
                 ; brackets (i.e., beginning with '['
                 ; and ending with ']')
                 ;
ifqdn          = 1*1023(domainpoint)
```

```
                    ;
                    ; a "domainpoint" is a UTF-8 encoded
                    ; Unicode code point that conforms to
                    ; RFC 5890
                    ;
```

Note well that all code points and blocks not explicitly allowed in the PRECIS IdentifierClass 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.

Note also that common constructions such as "user@example.com" are allowed as simple user names when using software that conforms to this specification, as they were under [RFC4013].

2.2. Preparation

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

Each simplepart of a simple user name MUST conform to the definition of the PRECIS IdentifierClass provided in [FRAMEWORK], where the normalization, casemapping, and directionality rules are as described below.

1. Unicode Normalization Form C (NFC) MUST be applied to all characters.

2. Uppercase and titlecase characters MUST be mapped to their lowercase equivalents.

3. Additional mappings MAY be applied, such as those defined in [I-D.ietf-precis-mappings].

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

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 [UNICODE], encoded using UTF-8 [RFC3629], and conformant to the PRECIS FreeformClass.

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 PRECIS "FreeformClass"
                 ;
```

Note well that all code points and blocks not explicitly allowed in the PRECIS FreeformClass 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.

3.2. Preparation

A password MUST NOT be zero bytes in length. This rule is to be enforced after any normalization and 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 FreeformClass.

With regard to directionality, the "Bidi Rule" (defined in [RFC5893]) and similar rules are unnecessary and inapplicable to passwords, since they can reduce the range of characters that are allowed in a string and therefore reduce the amount of entropy that is possible in a password. Furthermore, such rules are intended to minimize the possibility that the same string will be displayed differently on a system set for right-to-left display and a system set for left-to-right display; however, passwords are typically not displayed at all and are rarely meant to be interoperable across different systems in the way that non-secret strings like domain names and user names are.

4. Migration

The rules defined in this specification differ slightly from those defined by the SASLprep specification [RFC4013]. The following sections describe these differences, along with their implications for migration, in more detail.

4.1. User Names

Deployments that currently use SASLprep for handling user names might need to scrub existing data when migrating to use of the rules defined in this specification. In particular:

- SASLprep specified the use of Unicode Normalization Form KC (NFKC), whereas this usage of the PRECIS IdentifierClass 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 IdentifierClass entirely disallows Unicode code points with compatibility equivalents (i.e., during comparison NFKC is more "aggressive" about finding matches than is NFC). A few examples might suffice to indicate the nature of the problem: (1) U+017F LATIN SMALL LETTER LONG S is compatibility equivalent to U+0073 LATIN SMALL LETTER S (2) U+2163 ROMAN NUMERAL FOUR is compatibility equivalent to U+0049 LATIN CAPITAL LETTER I and U+0056 LATIN CAPITAL LETTER V (3) U+FB01 LATIN SMALL LIGATURE FI is compatibility equivalent to U+0066 LATIN SMALL LETTER F and U+0069 LATIN SMALL LETTER I. Under SASLprep, the use of NFKC also handled the mapping of fullwidth and halfwidth code points to their decomposition equivalents (see [I-D.ietf-precis-mappings]). Although it is expected that code points with compatibility equivalents are rare in existing user names, for migration purposes deployments might want to search their database of user names for Unicode code points with compatibility equivalents and map those code points to their compatibility equivalents.

- SASLprep mapped non-ASCII spaces to ASCII space (U+0020), whereas the PRECIS IdentifierClass entirely disallows non-ASCII spaces. The non-ASCII space characters are U+00A0 NO-BREAK SPACE, U+1680 OGHAM SPACE MARK, U+180E MONGOLIAN VOWEL SEPARATOR, U+2000 EN QUAD through U+200A HAIR SPACE, U+202F NARROW NO-BREAK SPACE, U+205F MEDIUM MATHEMATICAL SPACE, and U+3000 IDEOGRAPHIC SPACE. For migration purposes, deployments might want to convert non-ASCII space characters to ASCII space in simple user names.

- SASLprep mapped the "characters commonly mapped to nothing" from Appendix B.1 of [RFC3454]) to nothing, whereas the PRECIS IdentifierClass entirely disallows most of these 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 does not have a derived property of Default_Ignorable_Code_Point in Unicode 6.1). For migration purposes, deployments might want to remove code

points contained in the PRECIS "M" category from simple user names.

- SASLprep allowed uppercase and titlecase characters, whereas this usage of the PRECIS IdentifierClass 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.

4.2. Passwords

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:

- SASLprep specified the use of Unicode Normalization Form KC (NFKC), whereas this usage of the PRECIS FreeformClass 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 has the security benefit of probably resulting in fewer false positives when comparing passwords. A few examples might suffice to indicate the nature of the problem: (1) U+017F LATIN SMALL LETTER LONG S is compatibility equivalent to U+0073 LATIN SMALL LETTER S (2) U+2163 ROMAN NUMERAL FOUR is compatibility equivalent to U+0049 LATIN CAPITAL LETTER I and U+0056 LATIN CAPITAL LETTER V (3) U+FB01 LATIN SMALL LIGATURE FI is compatibility equivalent to U+0066 LATIN SMALL LETTER F and U+0069 LATIN SMALL LETTER I. Under SASLprep, the use of NFKC also handled the mapping of fullwidth and halfwidth code points to their decomposition equivalents (see [I-D.ietf-precis-mappings]). Although it is expected that code points with compatibility equivalents are rare in existing passwords, some passwords that matched when SASLprep was used might no longer work when the rules in this specification are applied.

- SASLprep mapped the "characters commonly mapped to nothing" from Appendix B.1 of [RFC3454]) to nothing, whereas the PRECIS FreeformClass 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). In practice, this change will probably have no effect on comparison, but user-oriented software might reject such code points instead of ignoring them during password preparation.

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 "IdentifierClass" and "FreeformClass" base string classes used in this document for simple 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 IdentifierClass

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

Applicability: Usernames in SASL and Kerberos.

Base Class: IdentifierClass.

Subclass: No.

Replaces: The SASLprep profile of Stringprep.

Normalization: NFC.

Casemapping: Map uppercase and titlecase characters to lowercase.

Additional Mappings: None.

Directionality: The "Bidi Rule" defined in RFC 5893 applies.

Specification: RFC XXXX. [Note to RFC Editor: please change XXXX to the number issued for this specification.]

6.2. Use of FreeformClass

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

Applicability: Passwords in SASL and Kerberos.

Base Class: FreeformClass

Subclass: No.

Replaces: The SASLprep profile of Stringprep.

Normalization: NFC.

Casemapping: None.

Additional Mappings: Map non-ASCII space characters to ASCII space.

Directionality: None.

Specification: RFC XXXX. [Note to RFC Editor: please change XXXX to the number issued for this specification.]

7. 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 under Section 4.)

8. References

8.1. Normative References

[FRAMEWORK] Saint-Andre, P. and M. Blanchet, "Precis Framework: Handling Internationalized Strings in Protocols", draft-ietf-precis-framework-07 (work in progress), March 2013.

[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/>.

8.2. Informative References

[I-D.ietf-precis-mappings] YONEYA, Y. and T. NEMOTO, "Mapping characters for PRECIS classes", draft-ietf-precis-mappings-01 (work in progress), December 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 4013.

- A single SASLprep algorithm was replaced by two separate algorithms: one for simple user names and another for passwords.
- The new preparation algorithms use PRECIS instead of a stringprep profile. The new algorithms work independenctly of Unicode versions.
- As recommended in the PRECIS framwork, changed the Unicode normalization form from NFKC to NFC.
- 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 implementation feedback. Thanks also to Marc Blanchet, Joe Hildebrand, Alan DeKok, Simon Josefsson, Jonathan Lennox, Matt Miller, Chris Newman, Pete Resnick, Andrew Sullivan, and Nico Williams for their input.

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