Network Working Group
Request for Comments: 4013
Category: Standards Track

K. Zeilenga
OpenLDAP Foundation

February 2005

SASLprep: Stringprep Profile for User Names and Passwords

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

This document describes how to prepare Unicode strings representing user names and passwords for comparison. The document defines the "SASLprep" profile of the "stringprep" algorithm to be used for both user names and passwords. This profile is intended to be used by Simple Authentication and Security Layer (SASL) mechanisms (such as PLAIN, CRAM-MD5, and DIGEST-MD5), as well as other protocols exchanging simple user names and/or passwords.

1. Introduction

The use of simple user names and passwords in authentication and authorization is pervasive on the Internet. To increase the likelihood that user name and password input and comparison work in ways that make sense for typical users throughout the world, this document defines rules for preparing internationalized user names and passwords for comparison. For simplicity and implementation ease, a single algorithm is defined for both user names and passwords.

The algorithm assumes all strings are comprised of characters from the Unicode [Unicode] character set.

This document defines the "SASLprep" profile of the "stringprep" algorithm [StringPrep].

The profile is designed for use in Simple Authentication and Security Layer ([SASL]) mechanisms, such as [PLAIN], [CRAM-MD5], and [DIGEST-MD5]. It may be applicable where simple user names and

passwords are used. This profile is not intended for use in preparing identity strings that are not simple user names (e.g., email addresses, domain names, distinguished names), or where identity or password strings that are not character data, or require different handling (e.g., case folding).

This document does not alter the technical specification of any existing protocols. Any specification that wishes to use the algorithm described in this document needs to explicitly incorporate this document and provide precise details as to where and how this algorithm is used by implementations of that specification.

2. The SASLprep Profile

This section defines the "SASLprep" profile of the "stringprep" algorithm [StringPrep]. This profile is intended for use in preparing strings representing simple user names and passwords.

This profile uses Unicode 3.2 [Unicode].

Character names in this document use the notation for code points and names from the Unicode Standard [Unicode]. For example, the letter "a" may be represented as either <U+0061> or <LATIN SMALL LETTER A>. In the lists of mappings and the 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 "[CONTROL CHARACTERS]") and do not come from the standard.

Note: A glossary of terms used in Unicode can be found in [Glossary]. Information on the Unicode character encoding model can be found in [CharModel].

2.1. Mapping

This profile specifies:

- non-ASCII space characters [StringPrep, C.1.2] that can be mapped to SPACE (U+0020), and

- the "commonly mapped to nothing" characters [StringPrep, B.1] that can be mapped to nothing.

2.2. Normalization

This profile specifies using Unicode normalization form KC, as described in Section 4 of [StringPrep].

2.3. Prohibited Output

This profile specifies the following characters as prohibited input:

- Non-ASCII space characters [StringPrep, C.1.2]
- ASCII control characters [StringPrep, C.2.1]
- Non-ASCII control characters [StringPrep, C.2.2]
- Private Use characters [StringPrep, C.3]
- Non-character code points [StringPrep, C.4]
- Surrogate code points [StringPrep, C.5]
- Inappropriate for plain text characters [StringPrep, C.6]
- Inappropriate for canonical representation characters [StringPrep, C.7]
- Change display properties or deprecated characters [StringPrep, C.8]
- Tagging characters [StringPrep, C.9]

2.4. Bidirectional Characters

This profile specifies checking bidirectional strings as described in [StringPrep, Section 6].

2.5. Unassigned Code Points

This profile specifies the [StringPrep, A.1] table as its list of unassigned code points.

3. Examples

The following table provides examples of how various character data is transformed by the SASLprep string preparation algorithm

#	Input	Output	Comments
1	I<U+00AD>X	IX	SOFT HYPHEN mapped to nothing
2	user	user	no transformation
3	USER	USER	case preserved, will not match #2
4	<U+00AA>	a	output is NFKC, input in ISO 8859-1
5	<U+2168>	IX	output is NFKC, will match #1
6	<U+0007>		Error - prohibited character
7	<U+0627><U+0031>		Error - bidirectional check

4. Security Considerations

This profile is intended to prepare simple user name and password strings for comparison or use in cryptographic functions (e.g., message digests). The preparation algorithm was specifically designed such that its output is canonical, and it is well-formed.

However, due to an anomaly [PR29] in the specification of Unicode normalization, canonical equivalence is not guaranteed for a select few character sequences. These sequences, however, do not appear in well-formed text. This specification was published despite this known technical problem. It is expected that this specification will be revised before further progression on the Standards Track (after [Unicode] and/or [StringPrep] specifications have been updated to address this problem).

It is not intended for preparing identity strings that are not simple user names (e.g., distinguished names, domain names), nor is the profile intended for use of simple user names that require different handling (such as case folding). Protocols (or applications of those protocols) that have application-specific identity forms and/or comparison algorithms should use mechanisms specifically designed for these forms and algorithms.

Application of string preparation may have an impact upon the feasibility of brute force and dictionary attacks. While the number of possible prepared strings is less than the number of possible Unicode strings, the number of usable names and passwords is greater than as if only ASCII was used. Though SASLprep eliminates some Unicode code point sequences as possible prepared strings, that elimination generally makes the (canonical) output forms practicable and prohibits nonsensical inputs.

User names and passwords should be protected from eavesdropping.

General "stringprep" and Unicode security considerations apply. Both are discussed in [StringPrep].

5. IANA Considerations

This document details the "SASLprep" profile of the [StringPrep] protocol. This profile has been registered in the stringprep profile registry.

Name of this profile: SASLprep
RFC in which the profile is defined: RFC 4013
Indicator whether or not this is the newest version of the profile: This is the first version of the SASPprep profile.

6. Acknowledgement

This document borrows text from "Preparation of Internationalized Strings (’stringprep’)" and "Nameprep: A Stringprep Profile for Internationalized Domain Names", both by Paul Hoffman and Marc Blanchet. This document is a product of the IETF SASL WG.

7. Normative References

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

[Unicode] 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/reports/tr27/) and by the "Unicode Standard Annex #28: Unicode 3.2" (http://www.unicode.org/reports/tr28/).

8. Informative References

[Glossary] The Unicode Consortium, "Unicode Glossary", <http://www.unicode.org/glossary/>.

[CharModel] Whistler, K. and M. Davis, "Unicode Technical Report #17, Character Encoding Model", UTR17, <http://www.unicode.org/unicode/reports/tr17/>, August 2000.

[SASL] Melnikov, A., Ed., "Simple Authentication and Security Layer (SASL)", Work in Progress.

[CRAM-MD5] Nerenberg, L., "The CRAM-MD5 SASL Mechanism", Work in Progress.

[DIGEST-MD5] Leach, P., Newman, C., and A. Melnikov, "Using Digest Authentication as a SASL Mechanism", Work in Progress.

[PLAIN] Zeilenga, K., Ed., "The Plain SASL Mechanism", Work in Progress.

[PR29] "Public Review Issue #29: Normalization Issue", <http://www.unicode.org/review/pr-29.html>, February 2004.

Author’s Address

Kurt D. Zeilenga
OpenLDAP Foundation

EMail: Kurt@OpenLDAP.org

Full Copyright Statement

Copyright (C) The Internet Society (2005).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the IETF’s procedures with respect to rights in IETF Documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.