Internet Engineering Task Force (IETF)
Request for Comments: 9598
Obsoletes: 8398
Updates: 5280
Category: Standards Track
ISSN: 2070-1721

A. Melnikov
Isode Ltd
W. Chuang
Google, Inc.
C. Bonnell
DigiCert

May 2024

Internationalized Email Addresses in X.509 Certificates

Abstract

This document defines a new name form for inclusion in the otherName field of an X.509 Subject Alternative Name and Issuer Alternative Name extension that allows a certificate subject to be associated with an internationalized email address.

This document updates RFC 5280 and obsoletes RFC 8398.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9598.

Table of Contents

1. Introduction
2. Conventions Used in This Document
3. Name Definitions
4. IDNA2008
5. Matching of Internationalized Email Addresses in X.509 Certificates
6. Name Constraints in Path Validation
7. Security Considerations
8. Differences from RFC 8398
9. IANA Considerations
10. References

10.1. Normative References
10.2. Informative References
Appendix A. ASN.1 Module
Appendix B. Example of SmtpUTF8Mailbox
Acknowledgments
Authors' Addresses

1. Introduction

[RFC5280] defines the rfc822Name subjectAltName name type for representing email addresses as described in [RFC5321]. The syntax of rfc822Name is restricted to a subset of US-ASCII characters and thus can't be used to represent internationalized email addresses [RFC6531]. This document defines a new otherName variant to represent internationalized email addresses. In addition, this document requires all email address domains in X.509 certificates to conform to IDNA2008 [RFC5890].

This document obsoletes [RFC8398]. The primary motivation of this document is to simplify the encoding of domain labels found in the domain part of internationalized email addresses. In particular, [RFC8398] specifies that domain labels are conditionally encoded using either A-labels or U-labels. This specification simplifies encoding and processing of domain labels by mandating that the A-label representation be used in all cases.

2. Conventions Used in This Document

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Name Definitions

The GeneralName structure [RFC5280] supports many different name forms including otherName for extensibility. This section specifies the SmtpUTF8Mailbox name form of otherName so that internationalized email addresses can appear in the subjectAltName of a certificate, the issuerAltName of a certificate, or anywhere else that GeneralName is used.

```
id-on-SmtpUTF8Mailbox OBJECT IDENTIFIER ::= { id-on 9 }

SmtpUTF8Mailbox ::= UTF8String (SIZE (1..MAX))
-- SmtpUTF8Mailbox conforms to Mailbox as specified
-- in Section 3.3 of RFC 6531. Additionally, all domain
-- labels included in the SmtpUTF8Mailbox value are
-- encoded as LDH labels. In particular, domain labels
-- are not encoded as U-labels and instead are encoded
-- using their A-label representation.
```

When the subjectAltName (or issuerAltName) extension contains an internationalized email address with a non-ASCII Local-part, the address MUST be stored in the SmtpUTF8Mailbox name form of otherName. The format of SmtpUTF8Mailbox is a modified version of the internationalized Mailbox that was defined in Section 3.3 of [RFC6531], which was derived from Mailbox as defined in Section 4.1.2 of [RFC5321]. [RFC6531] defines the following ABNF rules for Mailbox whose parts are modified for internationalization: Local-part, Dot-string, Quoted-string, QcontentSMTP, Domain, and Atom. In particular, Local-part was updated to also support UTF8-non-ascii. UTF8-non-ascii was described by Section 3.1 of [RFC6532]. Also, domain was extended to support U-labels, as defined in [RFC5890].

This document further refines internationalized Mailbox ABNF rules as described in [RFC6531] and calls this SmtpUTF8Mailbox. In SmtpUTF8Mailbox, labels that include non-ASCII characters MUST be stored in A-label (rather than U-label) form [RFC5890]. This restriction reduces complexity for implementations of the certification path validation algorithm defined in Section 6 of [RFC5280]. In SmtpUTF8Mailbox, domain labels that solely use ASCII characters (meaning neither A- nor U-labels) SHALL use NR-LDH restrictions as specified by Section 2.3.1 of [RFC5890]. NR-LDH stands for "Non-Reserved Letters Digits Hyphen" and is the set of LDH labels that do not have "--" characters in the third and forth character positions, which excludes "tagged domain names" such as A-labels. To facilitate octet-for-octet comparisons of SmtpUTF8Mailbox values, all NR-LDH and A-label labels that constitute the domain part SHALL only be encoded with lowercase letters. Consistent with the treatment of rfc822Name in [RFC5280], SmtpUTF8Mailbox is an envelope Mailbox and has no phrase (such as a common name) before it, has no comment (text surrounded in parentheses) after it, and is not surrounded by "<" and ">" characters.

Due to name constraint compatibility reasons described in Section 6, SmtpUTF8Mailbox subjectAltName MUST NOT be used unless the Local-part of the email address contains non-ASCII characters. When the Local-part is ASCII, rfc822Name subjectAltName MUST be used instead of SmtpUTF8Mailbox. This is compatible with legacy software that supports only rfc822Name (and not SmtpUTF8Mailbox). The appropriate usage of rfc822Name and SmtpUTF8Mailbox is summarized in Table 1 below.

SmtpUTF8Mailbox is encoded as UTF8String. The UTF8String encoding MUST NOT contain a Byte Order Mark (BOM) [RFC3629] to aid consistency across implementations, particularly for comparison.

Local-part char	subjectAltName
ASCII-only	rfc822Name
non-ASCII	SmtpUTF8Mailbox

Table 1: Email Address Formatting

Non-ASCII Local-part values may additionally include ASCII characters.

4. IDNA2008

To facilitate comparison between email addresses, all email address domains in X.509 certificates MUST conform to IDNA2008 [RFC5890] (and avoid any "mappings" mentioned in that document). Use of non-conforming email address domains introduces the possibility of conversion errors between alternate forms. This applies to SmtpUTF8Mailbox and rfc822Name in subjectAltName, issuerAltName, and anywhere else that these are used.

5. Matching of Internationalized Email Addresses in X.509 Certificates

Equivalence comparisons with SmtpUTF8Mailbox consist of a domain part step and a Local-part step. The comparison form for Local-parts is always UTF-8. The comparison form for domain parts is always performed with the LDH label ([RFC5890]) encoding of the relevant domain labels. The comparison of LDH labels in domain parts reduces complexity for implementations of the certification path validation algorithm as defined in Section 6 of [RFC5280] by obviating the need

to convert domain labels to their Unicode representation.

Comparison of two SmtpUTF8Mailboxes is straightforward with no setup work needed. They are considered equivalent if there is an exact octet-for-octet match.

Comparison of an SmtpUTF8Mailbox and rfc822Name will always fail. SmtpUTF8Mailbox values SHALL contain a Local-part that includes one or more non-ASCII characters, while rfc822Names only includes ASCII characters (including the Local-part). Thus, an SmtpUTF8Mailbox and rfc822Name will never match.

Comparison of SmtpUTF8Mailbox values with internationalized email addresses from other sources (such as received email messages, user input, etc.) requires additional setup steps for domain part and Local-part. The initial preparation for the email address to compare with the SmtpUTF8Mailbox value is to remove any phrases, comments, and "<" or ">" characters.

For the setup of the domain part, the following conversions SHALL be performed:

1. Convert all labels that constitute the domain part that include non-ASCII characters to A-labels, if not already in that form.

    a. Detect all U-labels present within the domain part using Section 5.1 of [RFC5891].

    b. Transform all detected U-labels (Unicode) to A-labels (ASCII) as specified in Section 5.5 of [RFC5891].

2. Convert all uppercase letters found within the NR-LDH and A-label labels that constitute the domain part to lowercase letters.

For the setup of the Local-part, the Local-part MUST be verified to conform to the requirements of [RFC6530] and [RFC6531], including being a string in UTF-8 form. In particular, the Local- part MUST NOT be transformed in any way, such as by doing case folding or normalization of any kind. The Local-part of an internationalized email address is already in UTF-8. Once setup is complete, they are again compared octet for octet.

To summarize non-normatively, the comparison steps, including setup, are:

1. If the domain contains U-labels, transform them to A-labels.

2. If any NR-LDH or A-label domain label in the domain part contains uppercase letters, lowercase them.

3. Compare strings octet for octet for equivalence.

This specification expressly does not define any wildcard characters, and SmtpUTF8Mailbox comparison implementations MUST NOT interpret any characters as wildcards. Instead, to specify multiple email addresses through SmtpUTF8Mailbox, the certificate MUST use multiple subjectAltNames or issuerAltNames to explicitly carry any additional email addresses.

6. Name Constraints in Path Validation

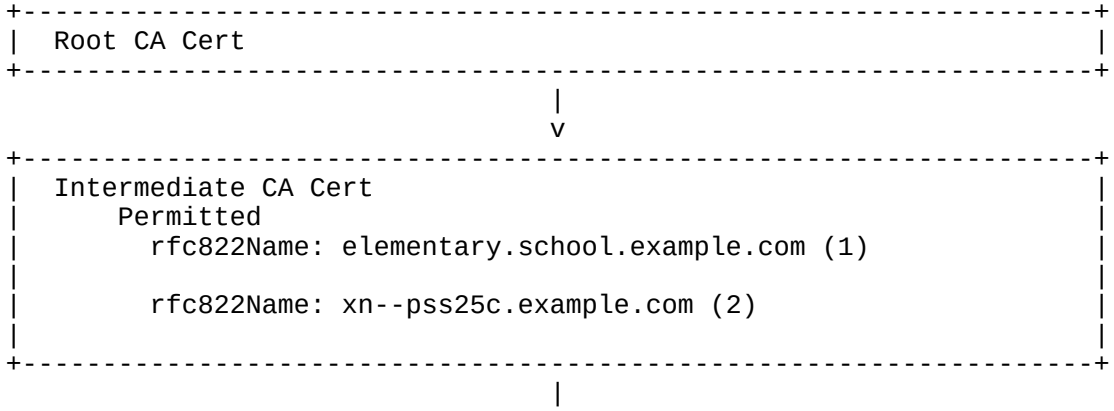
This section updates Section 4.2.1.10 of [RFC5280] to extend rfc822Name name constraints to SmtpUTF8Mailbox subjectAltNames. SmtpUTF8Mailbox-aware path validators will apply name constraint comparison to the subject distinguished name and both forms of subject alternative names, rfc822Name and SmtpUTF8Mailbox.

Both rfc822Name and SmtpUTF8Mailbox subject alternative names represent the same underlying email address namespace. Since legacy Certification Authorities (CAs) constrained to issue certificates for a specific set of domains would lack corresponding UTF-8 constraints, [RFC9549] updates, modifies, and extends rfc822Name name constraints defined in [RFC5280] to cover SmtpUTF8Mailbox subject alternative names. This ensures that the introduction of SmtpUTF8Mailbox does not violate existing name constraints. Since it is not valid to include non-ASCII UTF-8 characters in the Local-part of rfc822Name name constraints, and since name constraints that include a Local-part are rarely, if at all, used in practice, name constraints updated in [RFC9549] allow the forms that represent all addresses at a host, or all mailboxes in a domain and deprecates rfc822Name name constraints that represent a particular mailbox. That is, rfc822Name constraints with a Local-part SHOULD NOT be used.

Constraint comparison with SmtpUTF8Mailbox subjectAltName starts with the setup steps defined in Section 5. Setup converts the inputs of the comparison (which is one of a subject distinguished name, an rfc822Name, or an SmtpUTF8Mailbox subjectAltName, and one of an rfc822Name name constraint) to constraint comparison form. For both the name constraint and the subject, this will convert all A-labels and NR-LDH labels to lowercase. Strip the Local-part and "@" separator from each rfc822Name and SmtpUTF8Mailbox, which leaves just the domain part. After setup, follow the comparison steps defined in Section 4.2.1.10 of [RFC5280] as follows. If the resulting name constraint domain starts with a "." character, then for the name constraint to match, a suffix of the resulting subject alternative name domain MUST match the name constraint (including the leading ".") octet for octet. If the resulting name constraint domain does not start with a "." character, then for the name constraint to match, the entire resulting subject alternative name domain MUST match the name constraint octet for octet.

Certificate Authorities that wish to issue CA certificates with email address name constraints MUST use rfc822Name subject alternative names only. These MUST be IDNA2008-conformant names with no mappings and with non-ASCII domains encoded in A-labels only.

The name constraint requirement with an SmtpUTF8Mailbox subject alternative name is illustrated in the non-normative diagram in Figure 1. The first example (1) illustrates a permitted rfc822Name ASCII-only host name name constraint and the corresponding valid rfc822Name subjectAltName and SmtpUTF8Mailbox subjectAltName email addresses. The second example (2) illustrates a permitted rfc822Name host name name constraint with an A-label, and the corresponding valid rfc822Name subjectAltName and SmtpUTF8Mailbox subjectAltName email addresses. Note that an email address with an ASCII-only Local-part is encoded as rfc822Name despite also having Unicode present in the domain.



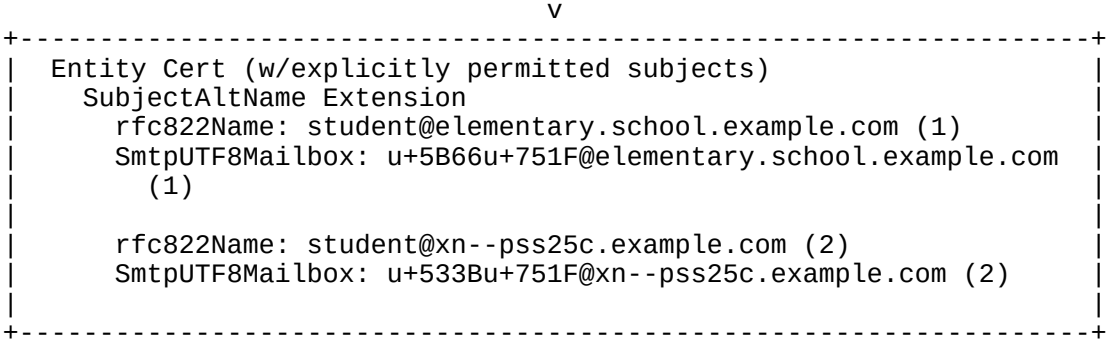


Figure 1: Name Constraints with SmtpUTF8Name and rfc822Name

7. Security Considerations

Use of SmtpUTF8Mailbox for certificate subjectAltName (and issuerAltName) will incur many of the same security considerations described in Section 8 of [RFC5280], but it introduces a new issue by permitting non-ASCII characters in the email address Local-part. This issue, as mentioned in Section 4.4 of [RFC5890] and in Section 4 of [RFC6532], is that use of Unicode introduces the risk of visually similar and identical characters that can be exploited to deceive the recipient. The former document references some means to mitigate against these attacks. See [WEBER] for more background on security issues with Unicode.

Additionally, it is possible to encode a string of Unicode user-perceived characters in multiple ways. While various Unicode normalization forms exist, [RFC6531] does not mandate the use of any such forms for the encoding of the Local-part. Thus, it may be possible to encode a Local-part value in multiple ways. To mitigate against attacks where different encodings are used by the mail system and the Certification Authority issues certificates containing SmtpUTF8Mailbox values, this specification requires an octet-for-octet comparison of the Local-part. However, requiring the use of binary comparison may raise interoperability concerns where the mail system employs one encoding and the Certification Authority employs another.

8. Differences from RFC 8398

This document obsoletes [RFC8398]. There are three major changes defined in this specification:

1. In all cases, domain labels in mail addresses SHALL be encoded as LDH labels. In particular, domain names SHALL NOT be encoded using U-Labels; instead, use A-Labels.

2. To accommodate the first change listed above, the mail address matching algorithm defined in Section 5 of [RFC8398] has been modified to only accept domain labels that are encoded using their A-label representation.

3. Additionally, the procedure to process rfc822Name name constraints as defined in Section 6 of [RFC8398] has been modified to only accept domain labels that are encoded using their A-label representation.

9. IANA Considerations

IANA has updated the reference for the id-mod-lamps-eai-addresses-2016 module in the "SMI Security for PKIX Module Identifier" (1.3.6.1.5.5.7.0) registry to refer to this document instead of [RFC8398].

IANA has updated the reference for the SmtpUTF8Mailbox otherName in the "SMI Security for PKIX Other Name Forms" (1.3.6.1.5.5.7.8) registry to refer to this document instead of [RFC8398].

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, DOI 10.17487/RFC3629, November 2003, <https://www.rfc-editor.org/info/rfc3629>.

[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, DOI 10.17487/RFC5280, May 2008, <https://www.rfc-editor.org/info/rfc5280>.

[RFC5321] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, DOI 10.17487/RFC5321, October 2008, <https://www.rfc-editor.org/info/rfc5321>.

[RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, DOI 10.17487/RFC5890, August 2010, <https://www.rfc-editor.org/info/rfc5890>.

[RFC5891] Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", RFC 5891, DOI 10.17487/RFC5891, August 2010, <https://www.rfc-editor.org/info/rfc5891>.

[RFC6530] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 6530, DOI 10.17487/RFC6530, February 2012, <https://www.rfc-editor.org/info/rfc6530>.

[RFC6531] Yao, J. and W. Mao, "SMTP Extension for Internationalized Email", RFC 6531, DOI 10.17487/RFC6531, February 2012, <https://www.rfc-editor.org/info/rfc6531>.

[RFC6532] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", RFC 6532, DOI 10.17487/RFC6532, February 2012, <https://www.rfc-editor.org/info/rfc6532>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8398] Melnikov, A., Ed. and W. Chuang, Ed., "Internationalized Email Addresses in X.509 Certificates", RFC 8398, DOI 10.17487/RFC8398, May 2018, <https://www.rfc-editor.org/info/rfc8398>.

[RFC9549] Housley, R., "Internationalization Updates to RFC 5280", RFC 9549, DOI 10.17487/RFC9549, March 2024, <https://www.rfc-editor.org/info/rfc9549>.

10.2. Informative References

[RFC5912] Hoffman, P. and J. Schaad, "New ASN.1 Modules for the

Public Key Infrastructure Using X.509 (PKIX)", RFC 5912, DOI 10.17487/RFC5912, June 2010, <https://www.rfc-editor.org/info/rfc5912>.

[WEBER] Weber, C., "Unraveling Unicode: A Bag of Tricks for Bug Hunting", July 2009, <https://www.lookout.net/files/Chris_Weber_Character%20Transformations%20v1.7_IUC33.pdf>.

Appendix A. ASN.1 Module

The following ASN.1 module normatively specifies the SmtpUTF8Mailbox structure. This specification uses the ASN.1 definitions from [RFC5912] with the 2002 ASN.1 notation used in that document. [RFC5912] updates normative documents using older ASN.1 notation.

```
LAMPS-EaiAddresses-2016
{ iso(1) identified-organization(3) dod(6)
  internet(1) security(5) mechanisms(5) pkix(7) id-mod(0)
  id-mod-lamps-eai-addresses-2016(92) }

DEFINITIONS IMPLICIT TAGS ::=
BEGIN

IMPORTS
OTHER-NAME
FROM PKIX1Implicit-2009
  { iso(1) identified-organization(3) dod(6) internet(1) security(5)
  mechanisms(5) pkix(7) id-mod(0) id-mod-pkix1-implicit-02(59) }

id-pkix
FROM PKIX1Explicit-2009
  { iso(1) identified-organization(3) dod(6) internet(1) security(5)
  mechanisms(5) pkix(7) id-mod(0) id-mod-pkix1-explicit-02(51) } ;

--
-- otherName carries additional name types for subjectAltName,
-- issuerAltName, and other uses of GeneralNames.
--

id-on OBJECT IDENTIFIER ::= { id-pkix 8 }

SmtpUtf8OtherNames OTHER-NAME ::= { on-SmtpUTF8Mailbox, ... }

on-SmtpUTF8Mailbox OTHER-NAME ::= {
    SmtpUTF8Mailbox IDENTIFIED BY id-on-SmtpUTF8Mailbox
}

id-on-SmtpUTF8Mailbox OBJECT IDENTIFIER ::= { id-on 9 }

SmtpUTF8Mailbox ::= UTF8String (SIZE (1..MAX))
-- SmtpUTF8Mailbox conforms to Mailbox as specified
-- in Section 3.3 of RFC 6531. Additionally, all domain
-- labels included in the SmtpUTF8Mailbox value are
-- encoded as LDH Labels. In particular, domain labels
-- are not encoded as U-Labels and instead are encoded
-- using their A-label representation.

END
```

Appendix B. Example of SmtpUTF8Mailbox

This non-normative example demonstrates using SmtpUTF8Mailbox as an otherName in GeneralName to encode the email address "u+533Bu+751F@xn--pss25c.example.com".

The hexadecimal DER encoding of the block is:

```
a02b0608 2b060105 05070809 a01f0c1d e58cbbe7 949f4078 6e2d2d70
73733235 632e6578 616d706c 652e636f 6d
```

The text decoding is:

```
0  43: [0] {
2   8:   OBJECT IDENTIFIER '1 3 6 1 5 5 7 8 9'
12  31:   [0] {
14  29:     UTF8String 'u+533Bu+751F@xn--pss25c.example.com'
      :     }
      :   }
```

The example was encoded using Google's "der-ascii" program and the above text decoding is an output of Peter Gutmann's "dumpasn1" program.

Acknowledgments

The authors thank David Benjamin for providing the motivation for this document. Additionally, the authors thank ric Vyncke, John Levine, Peter van Dijk, Rich Salz, Russ Housley, and Tim Hollebeek for their reviews and feedback, which meaningfully improved the document.

The authors also recognize and appreciate the following individuals for their contributions to [RFC8398]:

| Thank you to Magnus Nystrom for motivating this document. Thanks
| to Russ Housley, Nicolas Lidzborski, Laetitia Baudoin, Ryan
| Sleevi, Sean Leonard, Sean Turner, John Levine, and Patrik
| Falstrom for their feedback. Also special thanks to John Klensin
| for his valuable input on internationalization, Unicode, and ABNF
| formatting; to Jim Schaad for his help with the ASN.1 example and
| his helpful feedback; and especially to Viktor Dukhovni for
| helping us with name constraints and his many detailed document
| reviews.

Authors' Addresses

Alexey Melnikov
Isode Ltd
14 Castle Mews
Hampton, Middlesex
TW12 2NP
United Kingdom
Email: Alexey.Melnikov@isode.com

Wei Chuang
Google, Inc.
1600 Amphitheater Parkway
Mountain View, CA
United States of America
Email: weihaw@google.com

Corey Bonnell
DigiCert
Pittsburgh, PA
United States of America
Email: corey.bonnell@digicert.com