

Internet Engineering Task Force
Internet-Draft
Obsoletes: RFC5738 (if approved)
Intended status: Standards Track
Expires: January 12, 2012

P. Resnick, Ed.
Qualcomm Incorporated
C. Newman, Ed.
Oracle
S. Shen, Ed.
CNNIC
July 11, 2011

IMAP Support for UTF-8
draft-ietf-eai-5738bis-01

Abstract

This specification extends the Internet Message Access Protocol version 4rev1 (IMAP4rev1) to support UTF-8 encoded international characters in user names, mail addresses and message headers.

Status of This Memo

This Internet-Draft is submitted to IETF 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 January 12, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as

described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Conventions Used in this Document	3
3. UTF8=ACCEPT IMAP Capability	3
3.1. IMAP UTF-8 Quoted Strings	4
3.2. UTF8 Parameter to SELECT and EXAMINE	5
3.3. UTF-8 LIST and LSUB Responses	6
3.4. UTF-8 Interaction with IMAP4 LIST Command Extensions	6
3.4.1. UTF8 and UTF8ONLY LIST Selection Options	6
3.4.2. UTF8 LIST Return Option	7
4. UTF8=APPEND Capability	7
5. UTF8=USER Capability	8
6. UTF8=ALL Capability	8
7. UTF8=ONLY Capability	8
8. Up-Conversion Server Requirements	9
9. Issues with UTF-8 Header Mailstore	10
10. IANA Considerations	10
11. Security Considerations	11
12. References	12
12.1. Normative References	12
12.2. Informative References	13
Appendix A. Appendix A. Design Rationale	14
Appendix B. Appendix B. Examples Demonstrating Relationships between UTF8= Capabilities	15
Appendix C. Appendix C. Acknowledgments	15

1. Introduction

This specification extends IMAP4rev1 [RFC3501] to permit UTF-8 [RFC3629] in headers as described in "Internationalized Email Headers" [I-D.ietf-eai-rfc5335bis] . It also adds a mechanism to support mailbox names, login names, and passwords using the UTF-8 charset. This specification creates five new IMAP capabilities to allow servers to advertise these new extensions, along with two new IMAP LIST selection options and a new IMAP LIST return option.

This specification permits implementation of an IMAP server that hides mailboxes with internationalized email messages from IMAP clients that do not support this extension. Implementation of "Post-delivery Message Downgrading for Internationalized Email Messages" [popimap-downgrade] is necessary for an IMAP server to make mailboxes with internationalized email messages visible to IMAP clients that do not support this extension.

2. Conventions Used in this Document

The key words "MUST", "MUST NOT", "SHOULD", "SHOULD NOT", and "MAY" in this document are to be interpreted as defined in "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

The formal syntax uses the Augmented Backus-Naur Form (ABNF) [RFC5234] notation including the core rules defined in Appendix B of [RFC5234]. In addition, rules from IMAP4rev1 [RFC3501], UTF-8 [RFC3629], "Collected Extensions to IMAP4 ABNF" [RFC4466], and IMAP4 LIST Command Extensions [RFC5258] are also referenced.

In examples, "C:" and "S:" indicate lines sent by the client and server, respectively. If a single "C:" or "S:" label applies to multiple lines, then the line breaks between those lines are for editorial clarity only and are not part of the actual protocol exchange.

3. UTF8=ACCEPT IMAP Capability

The "UTF8=ACCEPT" capability indicates that the server supports UTF-8 quoted strings, the "UTF8" parameter to SELECT and EXAMINE, and UTF-8 responses from the LIST and LSUB commands.

A client MUST use the "ENABLE UTF8=ACCEPT" command (defined in [RFC5161]) to indicate to the server that the client accepts UTF-8 quoted-strings. The "ENABLE UTF8=ACCEPT" command MUST only be used in the authenticated state. (Note that the "UTF8=ONLY" capability described in Section 7 and the "UTF8=ALL" capability described in Section 6 imply the "UTF8=ACCEPT" capability. See additional

information in these sections.)

3.1. IMAP UTF-8 Quoted Strings

The IMAP4rev1 [RFC3501] base specification forbids the use of 8-bit characters in atoms or quoted strings. Thus, a UTF-8 string can only be sent as a literal. This can be inconvenient from a coding standpoint, and unless the server offers IMAP4 non-synchronizing literals [RFC2088], this requires an extra round trip for each UTF-8 string sent by the client. When the IMAP server advertises the "UTF8=ACCEPT" capability, it informs the client that it supports native UTF-8 quoted-strings with the following syntax:

```
string          =/ uQuoted
uQuoted         = "*" DQUOTE *uQUOTED-CHAR DQUOTE
                  ; referred as 'utf8-quote' in this document
DQUOTE          = <Defined in appendix B.1 of RFC 5234>
uQUOTED-CHAR    = QUOTED-CHAR / UTF8-2 / UTF8-3 / UTF8-4
UTF8-2          = <Defined in Section 4 of RFC3629>
UTF8-3          = <Defined in Section 4 of RFC3629>
UTF8-4          = <Defined in Section 4 of RFC3629>
```

When this quoting mechanism is used by the client (specifically an octet sequence beginning with "*" and ending with "), then the server MUST reject octet sequences with the high bit set that fail to comply with the formal syntax in [RFC3629] with a BAD response.

The IMAP server MUST NOT send utf8-quoted syntax to the client unless the client has indicated support for that syntax by using the "ENABLE UTF8=ACCEPT" command.

If the server advertises the "UTF8=ACCEPT" capability, the client MAY use utf8-quoted syntax with any IMAP argument that permits a string (including astring and nstring). However, if characters outside the US-ASCII repertoire are used in an inappropriate place, the results would be the same as if other syntactically valid but semantically invalid characters were used. For example, if the client includes UTF-8 characters in the user or password arguments (and the server has not advertised "UTF8=USER"), the LOGIN command will fail as it would with any other invalid user name or password. Specific cases where UTF-8 characters are permitted or not permitted are described

in the following paragraphs.

All IMAP servers that advertise the "UTF8=ACCEPT" capability SHOULD accept UTF-8 in mailbox names, and those that also support the "Mailbox International Naming Convention" described in RFC 3501, Section 5.1.3 MUST accept utf8-quoted mailbox names and convert them to the appropriate internal format. Mailbox names MUST comply with the Net-Unicode Definition (Section 2 of [RFC5198]) with the specific exception that they MUST NOT contain control characters (0000-001F, 0080-009F), delete (007F), line separator (2028), or paragraph separator (2029).

An IMAP client MUST NOT issue a SEARCH command that uses a mixture of utf8-quoted syntax and a SEARCH CHARSET other than UTF-8. If an IMAP server receives such a SEARCH command, it SHOULD reject the command with a BAD response (due to the conflicting charset labels).

3.2. UTF8 Parameter to SELECT and EXAMINE

The "UTF8=ACCEPT" capability also indicates that the server supports the "UTF8" parameter to SELECT and EXAMINE. When a mailbox is selected with the "UTF8" parameter, it alters the behavior of all IMAP commands related to message sizes, message headers, and MIME body headers so they refer to the message with UTF-8 headers. If the mailstore is not UTF-8 header native and the SELECT or EXAMINE command with UTF-8 header modifier succeeds, then the server MUST return results as if the mailstore were UTF-8 header native with upconversion requirements as described in Section 8. The server MAY reject the SELECT or EXAMINE command with the [NOT-UTF-8] response code, unless the "UTF8=ALL" or "UTF8=ONLY" capability is advertised.

Servers MAY include mailboxes that can only be selected or examined if the "UTF8" parameter is provided. However, such mailboxes MUST NOT be included in the output of an unextended LIST, LSUB, or equivalent command. If a client attempts to SELECT or EXAMINE such mailboxes without the "UTF8" parameter, the server MUST reject the command with a [UTF-8-ONLY] response code. As a result, such mailboxes will not be accessible by IMAP clients written prior to this specification and are discouraged unless the server advertises "UTF8=ONLY" or the server implements IMAP4 LIST Command Extensions [RFC5258].

utf8-select-param = "UTF8" ;; Conforms to select-param from RFC 4466

C: a SELECT newmailbox (UTF8)

S: ...

S: a OK SELECT completed

C: b FETCH 1 (SIZE ENVELOPE BODY)

S: ... UTF-8 header native results

S: b OK FETCH completed

C: c EXAMINE legacymailbox (UTF8)

S: c NO [NOT-UTF-8] Mailbox does not support UTF-8 access

C: d SELECT funky-new-mailbox

S: d NO [UTF-8-ONLY] Mailbox requires UTF-8 client

3.3. UTF-8 LIST and LSUB Responses

After an IMAP client successfully issues an "ENABLE UTF8=ACCEPT" command, the server MUST NOT return in LIST results any mailbox names to the client following the IMAP4 Mailbox International Naming Convention. Instead, the server MUST return any mailbox names with characters outside the US-ASCII repertoire using utf8-quoted syntax. (The IMAP4 Mailbox International Naming Convention has proved problematic in the past, so the desire is to make this syntax obsolete as quickly as possible.)

3.4. UTF-8 Interaction with IMAP4 LIST Command Extensions

When an IMAP server advertises both the "UTF8=ACCEPT" capability and the "LIST-EXTENDED" [RFC5258] capability, the server MUST support the LIST extensions described in this section.

3.4.1. UTF8 and UTF8ONLY LIST Selection Options

The "UTF8" LIST selection option tells the server to include mailboxes that only support UTF-8 headers in the output of the list

command. The "UTF8ONLY" LIST selection option tells the server to include all mailboxes that support UTF-8 headers and to exclude mailboxes that don't support UTF-8 headers. Note that "UTF8ONLY" implies "UTF8", so it is not necessary for the client to request both. Use of either selection option will also result in UTF-8 mailbox names in the result as described in Section 3.3 and implies the "UTF8" List return option described in Section 3.4.2.

3.4.2. UTF8 LIST Return Option

If the client supplies the "UTF8" LIST return option, then the server MUST include either the "\NoUTF8" or the "\UTF8Only" mailbox attribute as appropriate. The "\NoUTF8" mailbox attribute indicates that an attempt to SELECT or EXAMINE that mailbox with the "UTF8" parameter will fail with a [NOT-UTF-8] response code. The "\UTF8Only" mailbox attribute indicates that an attempt to SELECT or EXAMINE that mailbox without the "UTF8" parameter will fail with a [UTF-8-ONLY] response code. Note that computing this information may be expensive on some server implementations, so this return option should not be used unless necessary.

The ABNF [RFC5234] for these LIST extensions follows:

```
List-select-independent-opt =/ "UTF8"
                               ; List-select-independent-opt is defined in RFC 5258 Section 6

list-select-base-opt         =/ "UTF8ONLY"
                               ; list-select-base-opt is defined in RFC 5258 Section 6

return-option                 =/ "UTF8"
                               ; return-option is defined in RFC 5258 Section 6

mbx-list-oflag                =/ "\NoUTF8" / "\UTF8Only"
                               ; mbx-list-oflag is defined in RFC 3501 Section 9

resp-text-code                =/ "NOT-UTF-8" / "UTF-8-ONLY"
                               ; resp-text-code is defined in RFC 3501 Section 9
```

4. UTF8=APPEND Capability

If the "UTF8=APPEND" capability is advertised, then the server accepts UTF-8 headers in the APPEND command message argument. A client that sends a message with UTF-8 headers to the server MUST send them using the "UTF8" APPEND data extension. If the server also advertises the CATENATE capability (as specified in [RFC4469]), the client can use the same data extension to include such a message in a CATENATE message part. The ABNF for the APPEND data extension and CATENATE extension follows:

```
utf8-literal    = "UTF8" SP "(" literal8 ")"  
append-data     =/ utf8-literal  
cat-part        =/ utf8-literal
```

A server that advertises "UTF8=APPEND" MAY fail for \NotUTF8 mailboxes with a NOT-UTF-8 response code. If this command does not fail, it MAY follow the requirements of the IMAP base specification and [RFC5322] for message fetching. Mechanisms for 7-bit downgrading to help comply with the standards are discussed in [popimap-downgrade].

IMAP servers that do not advertise the "UTF8=APPEND" or "UTF8=ONLY" capability SHOULD reject an APPEND command that includes any 8-bit in the message headers with a "NO" response.

Note that the "UTF8=ONLY" capability described in Section 7 implies the "UTF8=APPEND" capability. See additional information in that section.

5. UTF8=USER Capability

If the "UTF8=USER" capability is advertised, that indicates the server accepts UTF-8 user names and passwords and applies SASLprep [RFC4013] to both arguments of the LOGIN command. The server MUST reject UTF-8 that fails to comply with the formal syntax in RFC 3629 [RFC3629] or if it encounters Unicode characters listed in Section 2.3 of SASLprep RFC 4013 [RFC4013].

6. UTF8=ALL Capability

The "UTF8=ALL" capability indicates all server mailboxes support UTF-8 headers. Specifically, SELECT and EXAMINE with the "UTF8" parameter will never fail with a [NOT-UTF-8] response code.

Note that the "UTF8=ONLY" capability described in Section 7 implies the "UTF8=ALL" capability. See additional information in that section.

Note that the "UTF8=ALL" capability implies the "UTF8=ACCEPT" capability.

7. UTF8=ONLY Capability

The "UTF8=ONLY" capability permits an IMAP server to advertise that it does not support the international mailbox name convention (modified UTF-7), and does not permit selection or examination of any

mailbox unless the "UTF8" parameter is provided. As this is an incompatible change to IMAP, a clear warning is necessary. IMAP clients that find implementation of the "UTF8=ONLY" capability problematic are encouraged to at least detect the "UTF8=ONLY" capability and provide an informative error message to the end-user.

The "UTF8=ONLY" capability implies the "UTF8=ACCEPT" capability, the "UTF8=ALL" capability, and the "UTF8=APPEND" capability. A server that advertises "UTF8=ONLY" need not advertise the three implicit capabilities.

8. Up-Conversion Server Requirements

When an IMAP4 server uses a traditional mailbox format that includes 7-bit headers and it chooses to permit access to that mailbox with the "UTF8" parameter, it **MUST** support minimal up-conversion as described in this section.

The server **MUST** support up-conversion of the following address header-fields in the message header: From, Sender, To, CC, Bcc, Resent-From, Resent-Sender, Resent-To, Resent-CC, Resent-Bcc, and Reply-To. This up-conversion **MUST** include address domains encoded according to Internationalizing Domain Names in Applications (IDNA) [RFC5890], and MIME header encoding [RFC2047] of display-names and any [RFC5322] comments.

The following charsets **MUST** be supported for up-conversion of MIME header encoding [RFC2047]: UTF-8, US-ASCII, ISO-8859-1, ISO-8859-2, ISO-8859-3, ISO-8859-4, ISO-8859-5, ISO-8859-6, ISO-8859-7, ISO-8859-8, ISO-8859-9, ISO-8859-10, ISO-8859-14, and ISO-8859-15. If the server supports other charsets in IMAP SEARCH or IMAP CONVERT [RFC5259], it **SHOULD** also support those charsets in this conversion.

Up-conversion of MIME header encoding of the following headers **MUST** also be implemented: Subject, Date ([RFC5322] comments only), Comments, Keywords, and Content-Description.

Server implementations also **SHOULD** up-convert all MIME body headers [RFC2045], **SHOULD** up-convert or remove the deprecated (and misused) "name" parameter [RFC1341] on Content-Type, and **MUST** up-convert the Content-Disposition [RFC2183] "filename" parameter, except when any of these are contained within a multipart/signed MIME body part (see below). These parameters can be encoded using the standard MIME parameter encoding [RFC2231] mechanism, or via non-standard use of MIME header encoding [RFC2047] in quoted strings.

The IMAP server **MUST NOT** perform up-conversion of headers and content of multipart/signed, as well as Original-Recipient and Return-Path.

9. Issues with UTF-8 Header Mailstore

When an IMAP server uses a mailbox format that supports UTF-8 headers and it permits selection or examination of that mailbox without the "UTF8" parameter, it is the responsibility of the server to comply with the IMAP4rev1 base specification [RFC3501] and [RFC5322] with respect to all header information transmitted over the wire. Mechanisms for 7-bit downgrading to help comply with the standards are discussed in [popimap-downgrade].

An IMAP server with a mailbox that supports UTF-8 headers MUST comply with the protocol requirements implicit from Section 8. However, the code necessary for such compliance need not be part of the IMAP server itself in this case. For example, the minimal required up-conversion could be performed when a message is inserted into the IMAP-accessible mailbox.

10. IANA Considerations

This document adds five new capabilities ("UTF8=ACCEPT", "UTF8=USER", "UTF8=APPEND", "UTF8=ALL", and "UTF8=ONLY") to the IMAP4rev1 Capabilities registry [RFC3501].

This document adds two new IMAP4 list selection options and one new IMAP4 list return option.

1. LIST-EXTENDED option name: UTF8

LIST-EXTENDED option type: SELECTION

Implied return options(s): UTF8

LIST-EXTENDED option description: Causes the LIST response to include mailboxes that mandate the UTF8 SELECT/EXAMINE parameter.

Published specification: RFC 5738bis, Section 3.4.1

Security considerations: RFC 5738bis, Section 11

Intended usage: COMMON

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

2. LIST-EXTENDED option name: UTF8ONLY

LIST-EXTENDED option type: SELECTION

Implied return options(s): UTF8

LIST-EXTENDED option description: Causes the LIST response to include mailboxes that mandate the UTF8 SELECT/EXAMINE parameter and exclude mailboxes that do not support the UTF8 SELECT/EXAMINE parameter.

Published specification: RFC 5738bis, Section 3.4.1

Security considerations: RFC 5738bis, Section 11

Intended usage: COMMON

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

3. LIST-EXTENDED option name: UTF8

LIST-EXTENDED option type: RETURN

Implied return options(s): none

LIST-EXTENDED option description: Causes the LIST response to include \NoUTF8 and \UTF8Only mailbox attributes.

Published specification: RFC 5738bis, Section 3.4.1

Security considerations: RFC 5738bis, Section 11

Intended usage: COMMON

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

11. Security Considerations

The security considerations of UTF-8 [RFC3629] and SASLprep [RFC4013] apply to this specification, particularly with respect to use of UTF-8 in user names and passwords. Otherwise, this is not believed to alter the security considerations of IMAP4rev1.

[**]

This document does not address downgrading scenarios, the security issues are discussed in [popimap-downgrade]

12. References

12.1. Normative References

- [RFC1341] Borenstein, N. and N. Freed, "MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies", RFC 1341, June 1992.
- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2183] Troost, R., Dorner, S., and K. Moore, "Communicating Presentation Information in Internet Messages: The Content-Disposition Header Field", RFC 2183, August 1997.
- [RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.
- [RFC3501] Crispin, M., "INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1", RFC 3501, March 2003.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.
- [RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", RFC 4013, February 2005.

- [RFC4466] Melnikov, A. and C. Daboo, "Collected Extensions to IMAP4 ABNF", RFC 4466, April 2006.
- [RFC4469] Resnick, P., "Internet Message Access Protocol (IMAP) CATENATE Extension", RFC 4469, April 2006.
- [RFC5161] Gulbrandsen, A. and A. Melnikov, "The IMAP ENABLE Extension", RFC 5161, March 2008.
- [RFC5198] Klensin, J. and M. Padlipsky, "Unicode Format for Network Interchange", RFC 5198, March 2008.
- [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.
- [RFC5258] Leiba, B. and A. Melnikov, "Internet Message Access Protocol version 4 - LIST Command Extensions", RFC 5258, June 2008.
- [RFC5259] Melnikov, A. and P. Coates, "Internet Message Access Protocol - CONVERT Extension", RFC 5259, July 2008.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.
- [I-D.ietf-eai-rfc5335bis] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", draft-ietf-eai-rfc5335bis-11 (work in progress), July 2011.
- [RFC5738] Resnick, P. and C. Newman, "IMAP Support for UTF-8", RFC 5738, March 2010.
- [RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, August 2010.

12.2. Informative References

- [RFC2049] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Five: Conformance Criteria and Examples",

RFC 2049, November 1996.

- [RFC2088] Myers, J., "IMAP4 non-synchronizing literals", RFC 2088, January 1997.
- [RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.
- [I-D.ietf-eai-rfc5721bis] Gellens, R., Newman, C., Yao, J., and K. Fujiwara, "POP3 Support for UTF-8", draft-ietf-eai-rfc5721bis-01 (work in progress), March 2011.
- [popimap-downgrade] Fujiwara, K., "Post-delivery Message Downgrading for Internationalized Email Messages", draft-ietf-eai-popimap-downgrade-00 (work in progress), October 2010.

Appendix A. Appendix A. Design Rationale

This non-normative section discusses the reasons behind some of the design choices in the above specification.

The basic approach of advertising the ability to access a mailbox in UTF-8 mode is intended to permit graceful upgrade, including servers that support multiple mailbox formats. In particular, it would be undesirable to force conversion of an entire server mailstore to UTF-8 headers, so being able to phase-in support for new mailboxes and gradually migrate old mailboxes is permitted by this design.

"UTF8=USER" is optional because many identity systems are US-ASCII only, so it's helpful to inform the client up front that UTF-8 won't work.

The "UTF8=ONLY" mechanism simplifies diagnosis of interoperability problems when legacy support goes away. In the situation where backwards compatibility is broken anyway, just-send-UTF-8 IMAP has the advantage that it might work with some legacy clients. However, the difficulty of diagnosing interoperability problems caused by a just-send-UTF-8 IMAP mechanism is the reason the "UTF8=ONLY" capability mechanism was chosen.

The up-conversion requirements are designed to balance the desire to deprecate and eventually eliminate complicated encodings (like MIME header encodings) without creating a significant deployment burden for servers. As IMAP4 servers already require a MIME parser, this

includes additional server up-conversion requirements not present in POP3 Support for UTF-8 [I-D.draft-ietf-eai-rfc5721bis].

The set of mandatory charsets comes from two sources: MIME requirements [RFC2049] and IETF Policy on Character Sets [RFC2277]. Including a requirement to up-convert widely deployed encoded ideographic charsets to UTF-8 would be reasonable for most scenarios, but may require unacceptable table sizes for some embedded devices. The open-ended recommendation to support widely deployed charsets avoids the political ramifications of attempting to list such charsets. The authors believe market forces, existing open-source software, and public conversion tables are sufficient to deploy the appropriate charsets.

Appendix B. Appendix B. Examples Demonstrating Relationships between UTF8= Capabilities

UTF8=ACCEPT UTF8=USER UTF8=APPEND

UTF8=ACCEPT UTF8=ALL

UTF8=ALL ; Note, same as above

UTF8=ACCEPT UTF8=USER UTF8=APPEND UTF8=ALL UTF8=ONLY

UTF8=USER UTF8=ONLY ; Note, same as above

Appendix C. Appendix C. Acknowledgments

The authors wish to thank the participants of the EAI working group for their contributions to this document with particular thanks to Harald Alvestrand, David Black, Randall Gellens, Arnt Gulbrandsen, Kari Hurtta, John Klensin, Xiaodong Lee, Charles Lindsey, Alexey Melnikov, Subramanian Moonesamy, Shawn Steele, Daniel Taharlev, and Joseph Yee for their specific contributions to the discussion.

Authors' Addresses

Pete Resnick (editor)
Qualcomm Incorporated
5775 Morehouse Drive
San Diego, CA 92121-1714
US

Phone: +1 858 651 4478
EMail: presnick@qualcomm.com

Chris Newman (editor)
Oracle
800 Royal Oaks
Monrovia, CA 91016
USA

Phone:
EMail: chris.newman@oracle.com

Sean Shen (editor)
CNNIC
No.4 South 4th Zhongguancun Street
Beijing, 100190
China

Phone: +86 10-58813038
EMail: shenshuo@cnnic.cn

Email Address Internationalization
(EAI)
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2012

K. Fujiwara
JPRS
July 11, 2011

Post-delivery Message Downgrading for Internationalized Email Messages
draft-ietf-eai-popimap-downgrade-02.txt

Abstract

The Email Address Internationalization (UTF8SMTP) extension allows UTF-8 characters in mail header fields. POP and IMAP servers support internationalized email messages. If a POP/IMAP client does not support Email Address Internationalization, POP/IMAP servers cannot send Internationalized Email Headers to the client and cannot remove the message. To avoid the situation, this document describes a conversion mechanism for internationalized Email messages to be traditional message format.

Status of This Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on January 12, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the BSD License.

Table of Contents

1. Introduction	3
2. Terminology	4
3. Updating RFC 5322	4
4. New Header Fields Definition	5
4.1. Unknown Header Fields' Preservation Header Fields	5
5. Email Header Fields Downgrading	6
5.1. Downgrading Method for Each ABNF Element	6
5.1.1. RECEIVED Downgrading	7
5.1.2. UNSTRUCTURED Downgrading	7
5.1.3. WORD Downgrading	7
5.1.4. COMMENT Downgrading	7
5.1.5. MIME-VALUE Downgrading	7
5.1.6. DISPLAY-NAME Downgrading	7
5.1.7. GROUP Downgrading	7
5.1.8. MAILBOX Downgrading	8
5.1.9. ENCAPSULATION Downgrading	8
5.1.10. TYPED-ADDRESS Downgrading	8
5.2. Downgrading Method for Each Header Field	8
5.2.1. Address Header Fields That Contain <address>s	9
5.2.2. Address Header Fields with Typed Addresses	9
5.2.3. Downgrading Non-ASCII in Comments	9
5.2.4. Received Header Field	10
5.2.5. MIME Content Header Fields	10
5.2.6. Non-ASCII in <unstructured>	10
5.2.7. Non-ASCII in <phrase>	10
5.2.8. Other Header Fields	10
6. MIME Body-Part Header Field Downgrading	11
7. Security Considerations	11
8. Implementation Notes	12
8.1. RFC 2047 Encoding	12
9. IANA Considerations	12
10. Acknowledgements	13
11. References	13
11.1. Normative References	13
11.2. Informative References	14
Appendix A. Examples	14
A.1. Downgrading Example	14

1. Introduction

Traditional mail systems, which are defined by [RFC5322], allow ASCII characters in mail header field values. The UTF8SMTP extension ([I-D.ietf-eai-frmrwk-4952bis] and [I-D.ietf-eai-rfc5335bis] allows UTF-8 characters in mail header field values.

If a header field contains non-ASCII characters, POP/IMAP servers cannot send Internationalized Email Headers to the client and cannot remove the message. This message downgrading mechanism converts mail header fields to an all-ASCII representation. The POP/IMAP servers can use the downgrading mechanism and send the Internationalized Email message as a traditional form.

[I-D.ietf-eai-rfc5335bis] allows UTF-8 characters to be used in mail header fields and MIME header fields. The message downgrading mechanism specified here describes the conversion method from the internationalized email messages that are defined in [I-D.ietf-eai-frmrwk-4952bis], and [I-D.ietf-eai-rfc5335bis] to the traditional email messages defined in [RFC5322].

There is no good way to convert "From:" and "Sender:" header fields, the draft need to update [RFC5322] to allow empty "From:" and "Sender:" header fields and it is described in Section 3.

Message Downgrading may be implemented in POP server and IMAP server only.

This document tries to define the message downgrading process clearly.

Downgrading consists of the following four parts:

- o Updating RFC 5322
- o New header field definitions
- o Email header field downgrading
- o MIME header field downgrading

In Section 4 of this document, header fields starting with "Downgraded-" are introduced. They preserve the original header fields.

Email header field downgrading is described in Section 5. It generates ASCII-only header fields.

MIME header fields are expanded in [I-D.ietf-eai-rfc5335bis]. MIME header field downgrading is described in Section 6. It generates ASCII-only MIME header fields.

Displaying downgraded messages that originally contained internationalized header fields is out of scope of this document. A POP/IMAP client which does not support UTF8 extension does not know internationalized message format described in [I-D.ietf-eai-rfc5335bis].

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

All specialized terms used in this specification are defined in the Email Address Internationalization (EAI) overview [I-D.ietf-eai-frmwk-4952bis], in the mail message specifications [RFC5322], or in the MIME documents [RFC2045] [RFC2047] [RFC2183] [RFC2231]. The terms "ASCII address", "internationalized email address", "non-ASCII address", "il8mail address", "UTF8SMTP", "message", and "mailing list" are used with the definitions from [I-D.ietf-eai-frmwk-4952bis].

This document depends on [I-D.ietf-eai-rfc5335bis]. Key words used in those documents are used in this document, too.

The term "non-ASCII" refers to a UTF-8 string that contains at least one non-ASCII character.

A "UTF8SMTP message" is an email message expanded by [I-D.ietf-eai-rfc5335bis].

3. Updating RFC 5322

"From:" header field or "Sender:" header field may contain non-ASCII addresses in internationalized Email messages. These non-ASCII addresses are not allowed in [RFC5322]. The draft proposes that the pop/imap downgrading uses <group> syntax and encodes non-ASCII addresses into <display-name> with empty <group-list> described in Section 5.

This specification redefines "From:", "Sender:", "Resent-From:" and "Resent-Sender:" header fields defined in Section 3.6.2 and 3.6.6 of [RFC5322] to allow <group> in the header fields.

```
from           = "From:" address-list CRLF
resent-from    = "Resent-From:" address-list CRLF
sender         = "Sender:" address CRLF
resent-sender  = "Resent-Sender:" address CRLF
```

[[Note in Draft: There are still outstanding questions about the use of group syntax that the WG should resolve, or confirm that it is willing to live with and figure out how to describe in the document, at IETF 81. They include

1. RFC 5322 does not allow group syntax in From and Sender header fields. Existing MUAs may become very confused when they see group syntax in originator fields.
2. Use of group syntax in this way will essentially make it impossible to reply to a message.
3. "Reply-To:" header field allows the group syntax in [RFC5322]. Is it correct ?
4. The ABNF syntax here is not yet complete.
5. Should the document explicitly recommend the use of comments, possibly with encoded words, to document the original non-ASCII mailboxes?

Suggestion made to the WG for more in depth discussion.]]

4. New Header Fields Definition

New header fields starting with "Downgraded-" are defined here to preserve those mail header field values that contain UTF-8 characters. During downgrading, one new "Downgraded-" header field is added for each mail header field that cannot be passed as-is to a POP/IMAP client that does not support UTF8 extension. The original mail header field is removed or rewritten. Only those mail header fields that contain non-ASCII characters are affected. The result of this process is a message that is compliant with existing email specifications [RFC5322]. The original internationalized information can be retrieved by examining the "Downgraded-" header fields that were added.

4.1. Unknown Header Fields' Preservation Header Fields

The unknown header fields' preservation header fields are defined to encapsulate those original header fields that contain non-ASCII characters and are not otherwise provided for in this specification.

The encapsulation header field name is the concatenation of "Downgraded-" and the original name. The value field holds the original header field value.

The header field syntax is specified as follows:

```
fields      =/ unknown-downgraded-headers ":" unstructured CRLF
unknown-downgraded-headers = "Downgraded-" original-header-field-name
original-header-field-name = field-name
field-name  = 1*ftext
ftext       = %d33-57 /           ; Any character except
                %d59-126         ; controls, SP, and ":".
```

To encapsulate a header field in a "Downgraded-" header field:

1. Generate a new "Downgraded-" header field whose value is the original header field value.
2. Treat the generated header field content as if it were unstructured, and then apply [RFC2047] encoding with charset UTF-8 as necessary so the result is ASCII.
3. Remove the original header field.

5. Email Header Fields Downgrading

This section defines the conversion method to ASCII for each header field that may contain non-ASCII characters.

[I-D.ietf-eai-rfc5335bis] expands "Received:" header fields; [RFC5322] describes ABNF elements <mailbox>, <word>, <comment>, <unstructured>; [RFC2045] describes ABNF element <value>.

5.1. Downgrading Method for Each ABNF Element

Header field downgrading is defined below for each ABNF element. Downgrading an unknown header field is also defined as ENCAPSULATION downgrading. Converting the header field terminates when no non-ASCII characters remain in the header field.

5.1.1. RECEIVED Downgrading

If the header field name is "Received:" and the FOR clause contains a non-ASCII address, remove the FOR clause from the header field. Other parts (not counting <comment>s) should not contain non-ASCII values.

5.1.2. UNSTRUCTURED Downgrading

If the header field has an <unstructured> field that contains non-ASCII characters, apply [RFC2047] encoding with charset UTF-8.

5.1.3. WORD Downgrading

If the header field has any <word> fields that contain non-ASCII characters, apply [RFC2047] encoding with charset UTF-8.

5.1.4. COMMENT Downgrading

If the header field has any <comment> fields that contain non-ASCII characters, apply [RFC2047] encoding with charset UTF-8.

5.1.5. MIME-VALUE Downgrading

If the header field has any <value> elements defined by [RFC2045] and the elements contain non-ASCII characters, encode the <value> elements according to [RFC2231] with charset UTF-8 and leave the language information empty. If the <value> element is <quoted-string> and it contains <CFWS> outside the DQUOTE, remove the <CFWS> before this conversion.

5.1.6. DISPLAY-NAME Downgrading

If the header field has any <address> (<mailbox> or <group>) elements and they have <display-name> elements that contain non-ASCII characters, encode the <display-name> elements according to [RFC2047] with charset UTF-8. DISPLAY-NAME downgrading is the same algorithm as WORD downgrading.

5.1.7. GROUP Downgrading

<group> is defined in Section 3.4 of [RFC5322]. The <group> elements may contain <mailbox>s which contain non-ASCII addresses.

If the header field has any <group> elements which contain <mailbox> elements that contain non-ASCII addresses, rewrite each <group> element as

"Internationalized_Address_Removed" display-name ENCODED_WORD ";;" [CFWS]

where the <ENCODED_WORD> is the original <group-list> encoded according to [RFC2047].

5.1.8. MAILBOX Downgrading

The <mailbox> elements have no equivalent format for non-ASCII addresses. If the header field has any <mailbox> elements that contain non-ASCII characters, rewrite each <mailbox> element to ASCII-only format. The <mailbox> element that contains non-ASCII characters is one of two formats.

[Display-name] "<" Utf8-addr-spec ">"

Utf8-addr-spec

Rewrite both as:

[Display-name] "Internationalized Address " Encoded-word
" Removed;;"

where the <Encoded-word> is the original <Utf8-addr-spec> encoded according to [RFC2047].

5.1.9. ENCAPSULATION Downgrading

If the header field contains non-ASCII characters and is such that no rule is given above, encapsulate it in a "Downgraded-" header field as described in Section 4.1 as a last resort.

Applying this procedure to "Received:" header field is prohibited.

5.1.10. TYPED-ADDRESS Downgrading

If the header field contains <utf-8-type-addr> and the <utf-8-type-addr> contains raw non-ASCII characters, it is in utf-8-address form. Convert it to utf-8-addr-xtext form. Those forms are described in [I-D.ietf-eai-rfc5337bis-dsn]. COMMENT downgrading is also performed in this case. If the address type is unrecognized and the header field contains non-ASCII characters, then fall back to using ENCAPSULATION downgrading on the entire header field.

5.2. Downgrading Method for Each Header Field

Header fields are listed in [RFC4021]. This section describes the downgrading method for each header field.

If the whole mail header field does not contain non-ASCII characters, email header field downgrading is not required. Each header field's downgrading method is described below.

5.2.1. Address Header Fields That Contain <address>s

From:
Sender:
To:
Cc:
Bcc:
Reply-To:
Resent-From:
Resent-Sender:
Resent-To:
Resent-Cc:
Resent-Bcc:
Resent-Reply-To:
Return-Path:
Disposition-Notification-To:

If the header field contains <group> elements that contain non-ASCII addresses, perform COMMENT downgrading, DISPLAY-NAME downgrading, and GROUP downgrading.

If the header field contains <mailbox> elements that contain non-ASCII addresses, perform COMMENT downgrading, DISPLAY-NAME downgrading, and MAILBOX downgrading.

5.2.2. Address Header Fields with Typed Addresses

Original-Recipient:
Final-Recipient:

If the header field contains non-ASCII characters, perform TYPED-ADDRESS downgrading.

5.2.3. Downgrading Non-ASCII in Comments

Date:
Message-ID:
Resent-Message-ID:
In-Reply-To:
References:

Resent-Date:
Resent-Message-ID:
MIME-Version:
Content-ID:
Content-Transfer-Encoding:
Content-Language:
Accept-Language:
Auto-Submitted:

These header fields do not contain non-ASCII characters except in comments. If the header field contains UTF-8 characters in comments, perform COMMENT downgrading.

5.2.4. Received Header Field

Received:

Perform COMMENT downgrading and RECEIVED downgrading.

5.2.5. MIME Content Header Fields

Content-Type:
Content-Disposition:

Perform MIME-VALUE downgrading and COMMENT downgrading.

5.2.6. Non-ASCII in <unstructured>

Subject:
Comments:
Content-Description:

Perform UNSTRUCTURED downgrading.

5.2.7. Non-ASCII in <phrase>

Keywords:

Perform WORD downgrading.

5.2.8. Other Header Fields

For all other header fields that contain non-ASCII characters, are user-defined, and are missing from this document or future defined header fields, perform ENCAPSULATION downgrading.

If the software understands the header field's structure and a downgrading algorithm other than ENCAPSULATION is applicable, that software SHOULD use that algorithm; ENCAPSULATION downgrading is used as a last resort.

Mailing list header fields (those that start in "List-") are part of this category.

6. MIME Body-Part Header Field Downgrading

MIME body-part header fields may contain non-ASCII characters [I-D.ietf-eai-rfc5335bis]. This section defines the conversion method to ASCII-only header fields for each MIME header field that contains non-ASCII characters. Parse the message body's MIME structure at all levels and check each MIME header field to see whether it contains non-ASCII characters. If the header field contains non-ASCII characters in the header field value, the header field is a target of the MIME body-part header field's downgrading. Each MIME header field's downgrading method is described below. COMMENT downgrading, MIME-VALUE downgrading, and UNSTRUCTURED downgrading are described in Section 5.

Content-ID:

The "Content-ID:" header field does not contain non-ASCII characters except in comments. If the header field contains UTF-8 characters in comments, perform COMMENT downgrading.

Content-Type:

Content-Disposition: Perform MIME-VALUE downgrading and COMMENT downgrading.

Content-Description: Perform UNSTRUCTURED downgrading.

7. Security Considerations

A downgraded message's header fields contain ASCII characters only. But they still contain MIME-encapsulated header fields that contain non-ASCII UTF-8 characters. Furthermore, the body part may contain UTF-8 characters. Implementations parsing Internet messages need to accept UTF-8 body parts and UTF-8 header fields that are MIME-encoded. Thus, this document inherits the security considerations of MIME-encoded header fields ([RFC2047] and [RFC3629]).

Rewriting header fields increases the opportunities for undetected spoofing by malicious senders. However, rewritten header fields are preserved into Downgraded-* header fields, and parsing Downgraded-* header fields enables the detection of spoofing caused by

downgrading.

The techniques described here invalidate methods that depend on digital signatures over any part of the message, which includes the top-level header fields and body-part header fields. Depending on the specific message being downgraded, the following techniques are likely to break: DomainKeys Identified Mail (DKIM), and possibly S/MIME and Pretty Good Privacy (PGP). The two obvious mitigations are to stick to 7-bit transport when using these techniques (as most/all of them presently require) or to make sure to have UTF8SMTP end-to-end when needed.

While information in any email header field should usually be treated with some suspicion, current email systems commonly employ various mechanisms and protocols to make the information more trustworthy. Currently, information in the new Downgraded-* header fields is usually not inspected by these mechanisms, and may be even less trustworthy than the traditional header fields. Note that the Downgraded-* header fields could have been inserted with malicious intent (and with content unrelated to the traditional header fields).

See the "Security Considerations" section in [I-D.ietf-eai-frmwkr-4952bis] for more discussion.

8. Implementation Notes

8.1. RFC 2047 Encoding

While [RFC2047] has a specific algorithm to deal with whitespace in adjacent encoded words, there are a number of deployed implementations that fail to implement the algorithm correctly. As a result, whitespace behavior is somewhat unpredictable in practice when multiple encoded words are used. While RFC 5322 states that implementations SHOULD limit lines to not more than 78 characters, implementations MAY choose to allow overly long encoded words in order to work around faulty [RFC2047] implementations. Implementations that choose to do so SHOULD have an optional mechanism to limit line length to 78 characters.

9. IANA Considerations

IANA is requested to refuse registration of all field names that start with "Downgraded-". For unknown header fields, use the downgrading method described in Section 4.1 to avoid conflicts with existing IETF activity (Email Address Internationalization).

10. Acknowledgements

This document draws heavily from the experimental in-transit message downgrading procedure described in RFC 5504 [RFC5504]. The contribution of the co-author of that earlier document, Y. Yoneya, are gratefully acknowledged.

11. References

11.1. Normative References

- [I-D.ietf-eai-frmwrk-4952bis] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", draft-ietf-eai-frmwrk-4952bis-10 (work in progress), September 2010.
- [I-D.ietf-eai-rfc5335bis] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", draft-ietf-eai-rfc5335bis-11 (work in progress), July 2011.
- [I-D.ietf-eai-rfc5337bis-dsn] Hansen, T., Newman, C., and A. Melnikov, "Internationalized Delivery Status and Disposition Notifications", draft-ietf-eai-rfc5337bis-dsn-02 (work in progress), April 2011.
- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2183] Troost, R., Dorner, S., and K. Moore, "Communicating Presentation Information in Internet Messages: The Content-Disposition Header Field", RFC 2183, August 1997.

- [RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.
- [RFC4021] Klyne, G. and J. Palme, "Registration of Mail and MIME Header Fields", RFC 4021, March 2005.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.

11.2. Informative References

- [RFC5504] Fujiwara, K. and Y. Yoneya, "Downgrading Mechanism for Email Address Internationalization", RFC 5504, March 2009.

Appendix A. Examples

A.1. Downgrading Example

This appendix shows an message downgrading example. Consider a received mail message where:

- o The sender address is a non-ASCII address, "NON-ASCII-local@example.com". Its display-name is "DISPLAY-local".
- o The "To:" header field contains two non-ASCII addresses, "NON-ASCII-remote1@example.net" and "NON-ASCII-remote2@example.com" Its display-names are "DISPLAY-remote1" and "DISPLAY-remote2".
- o The "Cc:" header field contains a non-ASCII address, "NON-ASCII-remote3@example.org". Its display-name is "DISPLAY-remote3".
- o Four display names contain non-ASCII characters.
- o The Subject header field is "NON-ASCII-SUBJECT", which contains non-ASCII characters.

- o There is an unknown header field "X-Unknown-Header" which contains non-ASCII characters.

```
Return-Path: <NON-ASCII-local@example.com>
Received: from ... by ... for <NON-ASCII-remote1@example.net>
Received: from ... by ... for <NON-ASCII-remote1@example.net>
From: DISPLAY-local <NON-ASCII-local@example.com>
To: DISPLAY-remote1 <NON-ASCII-remote1@example.net>,
    DISPLAY-remote2 <NON-ASCII-remote2@example.com>
Cc: DISPLAY-remote3 <NON-ASCII-remote3@example.org>
Subject: NON-ASCII-SUBJECT
Date: DATE
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Content-Type: text/plain; charset="UTF-8"
Content-Transfer-Encoding: 8bit
X-Unknown-Header: NON-ASCII-CHARACTERS
```

MAIL_BODY

Figure 1: Received message in a mail drop

The downgraded message is shown in Figure 2. "Return-Path:", "From:", "To:" and "Cc:" header fields are rewritten. "X-Unknown-Header:" is encapsulated as "Downgraded-X-Unknown-Header:".

```
Return-Path: Internationalized address
  =?UTF-8?Q?NON-ASCII-local@example.com?= removed;;
Received: from ... by ...
Received: from ... by ...
From: =?UTF-8?Q?DISPLAY-local?= Internationalized address
  =?UTF-8?Q?NON-ASCII-local@example.com?= removed;;
To: =?UTF-8?Q?DISPLAY-remote1?= Internationalized address
  =?UTF-8?Q?NON-ASCII-remote1@example.net?= removed;;,
  =?UTF-8?Q?DISPLAY-remote2?= Internationalized address
  =?UTF-8?Q?NON-ASCII-remote2@example.com?= removed;;,
Cc: =?UTF-8?Q?DISPLAY-remote3?= Internationalized address
  =?UTF-8?Q?NON-ASCII-remote3@example.org?= removed;;
Subject: =?UTF-8?Q?NON-ASCII-SUBJECT?=
Date: DATE
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Content-Type: text/plain; charset="UTF-8"
Content-Transfer-Encoding: 8bit
Downgraded-X-Unknown-Header: =?UTF-8?Q?NON-ASCII-CHARACTERS?=

MAIL_BODY
```

Figure 2: Downgraded message

Author's Address

Kazunori Fujiwara
Japan Registry Services Co., Ltd.
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Chiyoda-ku, Tokyo 101-0065
Japan

Phone: +81 3 5215 8451
EMail: fujiwara@jprs.co.jp

Network Working Group
Internet-Draft
Obsoletes: 5336 (if approved)
Intended status: Standards Track
Expires: January 8, 2012

J. Yao
W. Mao
CNNIC
July 7, 2011

SMTP Extension for Internationalized Email
draft-ietf-eai-rfc5336bis-11.txt

Abstract

This document specifies an SMTP extension for transport and delivery of email messages with internationalized email addresses or header information.

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 January 8, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	4
1.1.	Role of This Specification	4
1.2.	Terminology	4
1.3.	Updates to Other Specifications	5
2.	Overview of Operation	5
3.	Mail Transport-Level Protocol	5
3.1.	Framework for the Internationalization Extension	5
3.2.	The UTF8SMTPbis Extension	6
3.3.	Extended Mailbox Address Syntax	7
3.4.	MAIL Command Parameter Usage	9
3.5.	Non-ASCII addresses and Reply-codes	9
3.6.	Body Parts and SMTP Extensions	10
3.7.	Additional ESMTP Changes and Clarifications	10
3.7.1.	The Initial SMTP Exchange	10
3.7.2.	Mail eXchangers	11
3.7.3.	Trace Information	11
3.7.4.	UTF-8 Strings in Replies	11
4.	IANA Considerations	13
5.	Security Considerations	15
6.	Acknowledgements	16
7.	Change History	16
7.1.	draft-yao-eai-rfc5336bis: Version 00	16
7.2.	draft-ietf-eai-rfc5336bis: Version 00	17
7.3.	draft-ietf-eai-rfc5336bis: Version 01	17
7.4.	draft-ietf-eai-rfc5336bis: Version 02	17
7.5.	draft-ietf-eai-rfc5336bis: Version 03	17
7.6.	draft-ietf-eai-rfc5336bis: Version 04	17
7.7.	draft-ietf-eai-rfc5336bis: Version 05	17
7.8.	draft-ietf-eai-rfc5336bis: Version 06	17
7.9.	draft-ietf-eai-rfc5336bis: Version 07	17
7.10.	draft-ietf-eai-rfc5336bis: Version 08	17
7.11.	draft-ietf-eai-rfc5336bis: Version 09	18
7.12.	draft-ietf-eai-rfc5336bis: Version 10	18
7.13.	draft-ietf-eai-rfc5336bis: Version 11	18
8.	References	18
8.1.	Normative References	18
8.2.	Informative References	19
	Authors' Addresses	20

1. Introduction

The Simple Mail Transfer Protocol [RFC5321] provides a negotiation mechanism about service extension by which SMTP clients can discover SMTP server capabilities and make decisions for further processing. This document uses this mechanism and specifies an SMTP extension to permit internationalized email addresses (see section 1.2) in the SMTP envelope, and Unicode characters encoded in UTF-8 [RFC3629] in the headers. An extended overview of the extension model for internationalized email addresses and the email header appears in [RFC4952bis], referred to as "the framework document" or just as "framework" elsewhere in this specification.

[[anchor1: Note in Draft and to RFC Editor: The keyword represented in this document by "UTF8SMTPbis" (and in the XML source byUTF8SMTPbis) is a placeholder. The actual keyword will be assigned when the standards track SMTP extension in this series [RFC5336bis-SMTP] is approved for publication and should be substituted here. This paragraph should be treated as normative reference to that SMTP extension draft, creating a reference hold until it is approved by the IESG. This paragraph should be removed before RFC publication.]]

1.1. Role of This Specification

The framework document [RFC4952bis] specifies the requirements for, and describes components of, full internationalization of electronic mail. A thorough understanding of the information in that document and in the base Internet email specifications [RFC5321] [RFC5322] is necessary to understand and implement this specification.

This document specifies an element of the email internationalization work, specifically the definition of an SMTP extension for internationalized email address transport delivery.

1.2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

The terms "UTF-8 string" or "UTF-8 character" are used to refer to Unicode characters encoded in UTF-8. All other specialized terms used in this specification are defined in the framework document or in the base Internet email specifications. In particular, the terms "ASCII address", "internationalized email address", "non-ASCII address", "UTF8SMTPbis", "internationalized message", and "message" are used in this document according to the definitions in the framework document.

Non-ASCII characters or strings referred in this document MUST be expressed in UTF-8, a standard Unicode Encoding Form.

This specification uses Augmented BNF (ABNF) rules [RFC5234]. Some basic rules in this document can be found from [RFC5234] or [RFC5321] under the same names.

1.3. Updates to Other Specifications

This specification extends some syntax rules defined in RFC 5321 and permits internationalized email address in the envelope, but it does not modify RFC 5321. It permits data formats defined in [RFC5335bis], but it does not modify RFC 5322. It does require that the 8BITMIME extension [RFC6152] be announced by the UTF8SMTPbis-aware SMTP server and used with "BODY=8BITMIME" by the UTF8SMTPbis-aware SMTP client, but it does not modify the 8BITMIME specification in any way.

2. Overview of Operation

This specification describes an optional extension to the email transport mechanism that permits non-ASCII characters in both the envelope and header fields of messages, which are encoded in UTF-8. The extension is identified with the token "UTF8SMTPbis".

The internationalized email headers specification [RFC5335bis] provides the details of email header features enabled by this extension

3. Mail Transport-Level Protocol

3.1. Framework for the Internationalization Extension

The following service extension is defined:

1. The name of the SMTP service extension is "Internationalized Email".
2. The EHLO keyword value associated with this extension is "UTF8SMTPbis".
3. No parameter values are defined for this EHLO keyword value. In order to permit future (although unanticipated) extensions, the EHLO response MUST NOT contain any parameters for this keyword. The UTF8SMTPbis-aware SMTP client MUST ignore any parameters if they appear for this keyword; that is, the UTF8SMTPbis-aware SMTP client MUST behave as if the parameters do not appear. If an SMTP server includes UTF8SMTPbis in its EHLO response, it MUST be fully compliant with this version of this specification.

4. One OPTIONAL parameter "UTF8SMTPbis" is added to the MAIL command. The parameter has no value. If this parameter is set in the MAIL command, it indicates that the SMTP client is UTF8SMTPbis-aware and asserts that the envelope includes the non-ASCII address or the message being sent is internationalized message or the message being sent needs the UTF8SMTPbis support.
5. The maximum length of a MAIL command line is increased by 13 characters by the possible addition of the UTF8SMTPbis parameter. [[anchor6: RFC Editor: the number '13' will be replaced by the new number (2 spaces + length of the new keyword supposed to replace "UTF8SMTPbis").]]
6. One OPTIONAL parameter "UTF8SMTPbis" is added to the VRFY and EXPN commands. The parameter UTF8SMTPbis has no value. The parameter indicates that the SMTP client can accept Unicode characters in UTF-8 encoding in replies from the VRFY and EXPN commands.
7. No additional SMTP verbs are defined by this extension.
8. Servers offering this extension MUST provide support for, and announce, the 8BITMIME extension [RFC6152].
9. The reverse-path and forward-path of the SMTP MAIL and RCPT commands are extended to allow Unicode characters encoded in UTF-8 in mailbox names (addresses).
10. The mail message body is extended as specified in [RFC5335bis].
11. The UTF8SMTPbis extension is valid on the submission port [RFC4409], and can be used with LMTP [RFC2033].

3.2. The UTF8SMTPbis Extension

An SMTP server that announces this UTF8SMTPbis extension MUST be prepared to accept a UTF-8 string [RFC3629] in any position in which RFC 5321 specifies that a <mailbox> can appear. Although the characters in the <local-part> are permitted to contain non-ASCII characters, actual parsing of the <local-part>, and the delimiters used, are unchanged from the base email specification [RFC5321]. Any domain names to be looked up in the DNS MUST allow for [RFC5890] behavior. When doing lookups, the UTF8SMTPbis-aware SMTP client or server MUST either use a Unicode aware DNS library, or transform it to A-label defined in [RFC5890].

An SMTP client that receives the UTF8SMTPbis extension keyword in response to the EHLO command MAY transmit mailbox names within SMTP commands as internationalized strings in UTF-8 form. It MAY send a UTF-8 header [RFC5335bis] (which may also include mailbox names in UTF-8). It MAY transmit the domain parts of mailbox names within SMTP commands or the message header as A-labels or U-labels [RFC5890]. The presence of the UTF8SMTPbis extension does not change RFC 5321 server relaying behaviors.

If the UTF8SMTPbis SMTP extension is not offered by the SMTP server, the UTF8SMTPbis-aware SMTP client MUST NOT transmit an internationalized email address and MUST NOT transmit a mail message containing internationalized mail headers as described in [RFC5335bis] at any level within its MIME structure [RFC2045] and [RFC2047]. (For this paragraph, the internationalized domain name in the form of A-labels as specified in IDNA definitions [RFC5890] is not considered to be "internationalized".) Instead, if an UTF8SMTPbis-aware SMTP client (UTF8SMTPbis-aware SMTP sender) attempts to transfer an internationalized message and encounters an SMTP server that does not support the extension, it MUST make one of the following three choices and the priority order is 1, 2 and 3.

1. It MAY either reject the message during the SMTP transaction or accept the message and then generate and transmit a notification of non-deliverability. Such notification MUST be done as specified in RFC 5321 [RFC5321], RFC 3464 [RFC3464], and the internationalized delivery status and disposition notifications specification [RFC5337bis].
2. If and only if the UTF8SMTPbis-aware SMTP client (sender) is a Message Submission Agent ("MSA") [RFC4409] [RFC5598], it may choose its own way to deal with this scenario according to the provisions of [RFC4409] or its future versions. But the detailed specification of this process and its results are outside the scope of this document.
3. It MAY find an alternate route to the destination that permits UTF8SMTPbis. That route MAY be discovered by trying alternate Mail eXchanger (MX) hosts (using preference rules as specified in RFC 5321) or using other means available to the UTF8SMTPbis-aware SMTP client.

This document applies when an UTF8SMTPbis-aware SMTP client or server supports the UTF8SMTPbis extension. For all other cases, and for addresses and messages that do not require an UTF8SMTPbis extension, UTF8SMTPbis-aware SMTP clients and servers do not change the behavior specified in [RFC5321].

3.3. Extended Mailbox Address Syntax

RFC 5321, section 4.1.2, defines the syntax of a <Mailbox> entirely in terms of ASCII characters. This document extends <Mailbox> to add support of non-ASCII characters.

The key changes made by this specification include:

- o In order to update the <Mailbox> to support the internationalized email address, the <Mailbox> ABNF rule will be imported from RFC 5321 directly, and other related rules are imported from RFC 5321, RFC 5234, RFC 5890 or RFC 5335bis, or are extended in this

document.

- o Extend the definition of <sub-domain> to permit both the RFC 5321 definition and a UTF-8 string in a DNS label that is conforming with IDNA definitions [RFC5890].
- o Extend the definition of <atext> to permit both the RFC 5321 definition and a UTF-8 string. That string MUST NOT contain any of the ASCII graphics or controls characters.

The following ABNF rules will be imported from RFC 5321, section 4.1.2 directly:

- o <Mailbox>
- o <Local-part>
- o <Dot-string>
- o <Quoted-string>
- o <QcontentSMTP>
- o <Domain>
- o <Atom>

The following ABNF rule will be imported from RFC 5335bis, section 4.1 directly:

- o <UTF8-non-ascii>

The following ABNF rule will be imported from RFC 5234, appendix B.1 directly:

- o <DQUOTE>

The following ABNF rule will be imported from RFC 5890, section 2.3.2.1 directly:

- o <U-label>

The following rules are extended in ABNF [RFC5234] as follows.

```
sub-domain  =/ U-label
; extend the definition of sub-domain in RFC5321, section 4.1.2

atext       =/ UTF8-non-ascii
; extend the definition of atext in RFC5321, section 4.1.2

quoted-pairSMTP =/ %d92 UTF8-non-ascii
; extend the definition of quoted-pairSMTP in RFC5321, section 4.1.2

qtextSMTP   =/ UTF8-non-ascii
; extend the definition of qtextSMTP in RFC5321, section 4.1.2
```

3.4. MAIL Command Parameter Usage

If the envelope or message being sent requires the capabilities of the UTF8SMTPbis extension, the UTF8SMTPbis-aware SMTP client **MUST** supply the UTF8SMTPbis parameter with the MAIL command. If this parameter is provided, it **MUST** have no value. If the UTF8SMTPbis-aware SMTP client is aware that neither the envelope nor the message being sent requires any of the UTF8SMTPbis extension capabilities, it **SHOULD NOT** supply the UTF8SMTPbis parameter with the MAIL command.

Because there is no guarantee that a next-hop SMTP server will support the UTF8SMTPbis extension, use of the UTF8SMTPbis extension always carries a risk of transmission failure. In fact, during the early stages of deployment for the UTF8SMTPbis extension, the risk will be quite high. Hence there is a distinct near-term advantage for ASCII-only messages to be sent without using this extension. The long-term advantage of casting ASCII [ASCII] characters(0x7f and below) as UTF-8 form is that it permits pure-Unicode environments.

This specification does not require that the UTF8SMTPbis-aware SMTP client inspect the message or otherwise go to extraordinary lengths to assure itself whether the UTF8SMTPbis extension is **REQUIRED** for the particular message.

3.5. Non-ASCII addresses and Reply-codes

An UTF8SMTPbis-aware SMTP client **MUST** only send an internationalized message to an SMTP server that supports UTF8SMTPbis. If the SMTP server does not support this option, then the UTF8SMTPbis-aware SMTP client has three choices according to section 3.2 of this specification.

The three-digit Reply-codes used in this section are based on their meanings as defined in RFC 5321.

When messages are rejected because the RCPT command requires an ASCII address, the reply-code 553 is returned with the meaning "mailbox name not allowed". When messages are rejected because the MAIL command requires an ASCII address, the reply-code 550 is returned with the meaning "mailbox unavailable". When the UTF8SMTPbis-aware SMTP server supports enhanced mail system status codes [RFC3463], reply-code "X.6.7" [RFC5248] is used, meaning that "non-ASCII addresses not permitted for that sender/recipient".

When messages are rejected for other reasons, the server follows the model of the base email specifications in RFC 5321; this extension does not change those circumstances or reply messages.

If the reply-code is issued after the final "." of the DATA command, the reply-code "554" is used with the meaning "Transaction failed". When the UTF8SMTPbis-aware SMTP server supports enhanced mail system status codes [RFC3463], reply-code "X.6.9" [RFC5248] is used, meaning that "UTF-8 header message can not be transmitted to one or more recipients, so the message MUST be rejected".

3.6. Body Parts and SMTP Extensions

The MAIL command parameter UTF8SMTPbis asserts that a message is an internationalized message or the message being sent needs the UTF8SMTPbis support. The message being sent via the MAIL command with the UTF8SMTPbis parameter has still a chance of that the message transmitted is not an internationalized message. An UTF8SMTPbis-aware SMTP client or server that requires accurate knowledge of whether a message is internationalized needs to parse all message header fields and MIME header fields [RFC2045] and [RFC2047] in the message body. However, this specification does not require that the UTF8SMTPbis-aware SMTP client or server inspects the message.

While this specification requires that UTF8SMTPbis-aware SMTP servers support the 8BITMIME extension [RFC6152] to ensure that servers have adequate handling capability for 8-bit data and to avoid a number of complex encoding problems, the use of internationalized email addresses obviously does not require non-ASCII body parts in the MIME message in RFC 2045 and RFC 2047. The UTF8SMTPbis extension MAY be used with the BODY=8BITMIME parameter [RFC6152] if that is appropriate given the body content or, with the BODY=BINARYMIME parameter, if the SMTP server advertises BINARYMIME [RFC3030] and that is appropriate.

3.7. Additional ESMTP Changes and Clarifications

The information carried in the mail transport process involves addresses ("mailboxes") and domain names in various contexts in addition to the MAIL and RCPT commands and extended alternatives to them. In general, the rule is that, when RFC 5321 specifies a mailbox, this SMTP extension requires UTF-8 form to be used for the entire string; when RFC 5321 specifies a domain name, the name SHOULD be in the form of A-label if this domain name is an internationalized domain name [RFC5890].

The following subsections list and discuss all of the relevant cases.

3.7.1. The Initial SMTP Exchange

When an SMTP connection is opened, the SMTP server sends a "greeting" response consisting of the 220 reply-code and some information. The

SMTP client then sends the EHLO command. Since the SMTP client cannot know whether the SMTP server supports UTF8SMTPbis until after it receives the response from EHLO, the UTF8SMTPbis-aware SMTP client MUST send only ASCII (LDH label or A-label [RFC5890]) domains in the EHLO command and that, if the UTF8SMTPbis-aware SMTP server provides domain names in the EHLO response, they MUST be in the form of LDH labels or A-labels.

3.7.2. Mail eXchangers

If multiple DNS MX records are used to specify multiple servers for a domain in section 5 of [RFC5321], it is strongly advised that all or none of them SHOULD support the UTF8SMTPbis extension. Otherwise, surprising rejections can happen during temporary or permanent failures, which users might perceive as serious reliability issues.

3.7.3. Trace Information

The trace information <Return-path-line>, <Time-stamp-line> and their related rules are defined in in section 4.4 of RFC 5321 [RFC5321]. This document updates <Mailbox> and <Domain> to support non-ASCII characters. When the UTF8SMTPbis extension is used, the 'Reverse-path' clause of the Return-path-line may include an internationalized domain name that uses the U-label form; The 'Stamp' clause of the Time-stamp-line may include an internationalized domain name that uses the U-label form.

If the messages that include trace fields are sent by an UTF8SMTPbis-aware SMTP client or relay server without the UTF8SMTPbis parameter at MAIL commands, trace field values must conform to RFC 5321 regardless of the SMTP server's capability.

If the messages that include trace fields are sent by an UTF8SMTPbis-aware SMTP client or relay server with the UTF8SMTPbis parameter at MAIL commands, trace field values SHOULD use the U-label form for the internationalized domain name.

The protocol value of the 'WITH' clause when this extension is used is one of the UTF8SMTPbis values specified in the "IANA Considerations" section of this document.

3.7.4. UTF-8 Strings in Replies

3.7.4.1. MAIL Command

If an SMTP client follows this specification and sends any MAIL commands containing the UTF8SMTPbis parameter, the UTF8SMTPbis-aware SMTP server is permitted to use UTF-8 characters in the email address

associated with 251 and 551 reply-codes, and the SMTP client MUST be able to accept and process them. If a given MAIL command does not include the UTF8SMTPbis parameter, the UTF8SMTPbis-aware SMTP server MUST NOT return a 251 or 551 response containing a non-ASCII mailbox. Instead, it MUST transform such responses into 250 or 550 responses that do not contain non-ASCII addresses.

3.7.4.2. VRFY and EXPN Commands and the UTF8SMTPbis Parameter

If the VRFY and EXPN commands are transmitted with the parameter "UTF8SMTPbis", it indicates the SMTP client can accept UTF-8 strings in replies to those commands. This parameter for the VRFY and EXPN commands SHOULD only be used after the SMTP client sees the EHLO response with the UTF8SMTPbis keyword. This allows the UTF8SMTPbis-aware SMTP server to use UTF-8 strings in mailbox names and full names that occur in replies without concern that the SMTP client might be confused by them. An SMTP client that conforms to this specification MUST accept and correctly process replies from the VRFY and EXPN commands that contain UTF-8 strings. However, the UTF8SMTPbis-aware SMTP server MUST NOT use UTF-8 strings in replies if the SMTP client does not specifically allow such replies by transmitting this parameter. Most replies do not require that a mailbox name be included in the returned text, and therefore UTF-8 string is not needed in them. Some replies, notably those resulting from successful execution of the VRFY and EXPN commands, do include the mailbox.

VERIFY (VRFY) and EXPAND (EXPN) command syntaxes are changed to:

```
vrfy = "VRFY" SP String
      [ SP "UTF8SMTPbis" ] CRLF
      ; String may include Non-ASCII characters

expn = "EXPN" SP String
      [ SP "UTF8SMTPbis" ] CRLF
      ; String may include Non-ASCII characters
```

The "UTF8SMTPbis" parameter does not use a value. If the reply to a VERIFY (VRFY) or EXPAND (EXPN) command requires UTF-8 string, but the SMTP client did not use the "UTF8SMTPbis" parameter, then the UTF8SMTPbis-aware SMTP server MUST use either the reply-code 252 or 550. Reply-code 252, defined in [RFC5321], means "Cannot VRFY user, but will accept the message and attempt the delivery". Reply-code 550, also defined in [RFC5321], means "Requested action not taken: mailbox unavailable". When the UTF8SMTPbis-aware SMTP server supports enhanced mail system status codes [RFC3463], the enhanced reply-code as specified below is used. Using the "UTF8SMTPbis"

parameter with a VERIFY (VRFY) or EXPAND (EXPN) command enables UTF-8 replies for that command only.

If a normal success response (i.e., 250) is returned, the response MAY include the full name of the user and MUST include the mailbox of the user. It MUST be in either of the following forms:

User Name <Mailbox>

- ; Mailbox is defined in section 3.3 of this document.
- ; User Name can contain non-ASCII characters.

Mailbox

- ; Mailbox is defined in section 3.3 of this document.

If the SMTP reply requires UTF-8 strings, but UTF-8 string is not allowed in the reply, and the UTF8SMTPbis-aware SMTP server supports enhanced mail system status codes [RFC3463], the enhanced reply-code is "X.6.8" [RFC5248], meaning "A reply containing a UTF-8 string is REQUIRED to show the mailbox name, but that form of response is not permitted by the SMTP client".

If the SMTP client does not support the UTF8SMTPbis extension, but receives a UTF-8 string in a reply, it may not be able to properly report the reply to the user, and some clients might crash. Internationalized messages in replies are only allowed in the commands under the situations described above. Under any other circumstances, UTF-8 string MUST NOT appear in the reply.

Although UTF-8 form is needed to represent email addresses in responses under the rules specified in this section, this extension does not permit the use of UTF-8 string for any other purposes. UTF8SMTPbis-aware SMTP servers MUST NOT include non-ASCII characters in replies except in the limited cases specifically permitted in this section.

4. IANA Considerations

IANA is requested to add a new value "UTF8SMTPbis" to the SMTP Service Extension subregistry of the Mail Parameters registry, according to the following data:

Keywords	Description	Reference
UTF8SMTPbis	Internationalized email address	[RFCXXXX]

This document updates the values to the SMTP Enhanced Status Code subregistry of the Mail Parameters registry, following the guidance in Sections 3.5 and 3.7.4.2 of this document, and being based on [RFC5248]. The registration data is as follows:

Code: X.6.7
Sample Text: non-ASCII addresses not permitted
for that sender/recipient
Associated basic status code: 550, 553
Description: This indicates the reception of a MAIL or RCPT
command that non-ASCII addresses are not permitted
Defined: RFC XXXX (Standard track)
Submitter: Jiankang YAO
Change controller: ima@ietf.org

Code: X.6.8
Sample Text: UTF-8 string reply is required,
but not permitted by the SMTP client
Associated basic status code: 252, 550, 553
Description: This indicates that a reply containing a UTF-8
string is required to show the mailbox name,
but that form of response is not
permitted by the SMTP client.
Defined: RFC XXXX (Standard track)
Submitter: Jiankang YAO
Change controller: ima@ietf.org

Code: X.6.9
Sample Text: UTF-8 header message can not be transferred
to one or more recipient so the message
must be rejected
Associated basic status code: 550
Description: This indicates that transaction failed
after the final "." of the DATA command.
Defined: RFC XXXX (Standard track)
Submitter: Jiankang YAO
Change controller: ima@ietf.org

Code: X.6.10
Description: This is a duplicate of X.6.8 and
is thus deprecated.

The following entries SHOULD be updated or added in the "Mail Transmission Types" registry under the Mail Parameters registry.

WITH protocol types	Description	Reference
UTF8SMTP	ESMTP with UTF8SMTP	[RFCXXXX]
UTF8SMTPA	ESMTP with UTF8SMTP and SMTP AUTH	[RFC4954] [RFCXXXX]
UTF8SMTPS	ESMTP with UTF8SMTP and STARTTLS	[RFC3207] [RFCXXXX]
UTF8SMTPSA	ESMTP with UTF8SMTP and both STARTTLS and SMTP AUTH	[RFC3207] [RFC4954] [RFCXXXX]
UTF8LMTP	LMTP with UTF8SMTP	[RFCXXXX]
UTF8LMTPA	LMTP with UTF8SMTP and SMTP AUTH	[RFC4954] [RFCXXXX]
UTF8LMTPS	LMTP with UTF8SMTP and STARTTLS	[RFC3207] [RFCXXXX]
UTF8LMTPSA	LMTP with UTF8SMTP and both STARTTLS and LMTP AUTH	[RFC3207] [RFC4954] [RFCXXXX]

5. Security Considerations

The extended security considerations discussion in the framework document [RFC4952bis] will be applied here.

More security considerations are discussed below:

Beyond the use inside the email global system (in SMTP envelopes and message headers), internationalized email addresses will also show up inside other cases, in particular:

- o the logging systems of SMTP transactions and other logs to monitor the email systems;
- o the trouble ticket systems used by Security Teams to manage security incidents, when an email address is involved;

In order to avoid problems that could cause loss of data, this will likely require extending these systems to support full UTF-8, or to require to provide an adequate mechanisms for mapping non-ASCII strings to ASCII.

Another security aspect to be considered is related to the ability by security team members to quickly understand, read and identify email addresses from the logs, when they are tracking an incident.

Mechanisms to automatically and quickly provide the origin or ownership of an internationalized email address SHALL be implemented for use also by log readers which cannot read easily non-ASCII information.

The SMTP commands VRFY and EXPN are sometimes used in SMTP transactions where there is no message to transfer (by tools used to take automated actions in case potential spam messages are identified). RFC 5321 section 3.5 and 7.3 give some detailed description of use and possible behaviours. Implementation of internationalized addresses can affect also logs and actions by these tools.

6. Acknowledgements

This document revised the [RFC5336] document based on the Email Address Internationalization (EAI) WG's discussion result. Many EAI WG members did some tests and implementations to move this document to the Standard Track document. Significant comments and suggestions were received from Xiaodong LEE, Nai-Wen Hsu, Yangwoo KO, Yoshiro YONEYA, and other members of the JET team and were incorporated into the specification. Additional important comments and suggestions, and often specific text, were contributed by many members of the WG and design team. Those contributions include material from John C Klensin, Charles Lindsey, Dave Crocker, Harald Tveit Alvestrand, Marcos Sanz, Chris Newman, Martin Duerst, Edmon Chung, Tony Finch, Kari Hurtta, Randall Gellens, Frank Ellermann, Alexey Melnikov, Pete Resnick, S. Moonesamy, Soobok Lee, Shawn Steele, Alfred Hoenes, Miguel Garcia, Magnus Westerlund, and Lars Eggert. Of course, none of the individuals are necessarily responsible for the combination of ideas represented here.

Thanks a lot to Dave Crocker for his comments and helping of ABNF refinement.

7. Change History

[[anchor14: RFC Editor: Please remove this section.]]

7.1. draft-yao-eai-rfc5336bis: Version 00

Applied errata suggested by Alfred Hoenes.

7.2. draft-ietf-eai-rfc5336bis: Version 00

Applied the changes suggested by the EAI new charter.

7.3. draft-ietf-eai-rfc5336bis: Version 01

Applied the changes suggested by 78 IETF EAI meeting.

7.4. draft-ietf-eai-rfc5336bis: Version 02

remove the appendix since rfc4952bis has added this material

improve the text

remove the text about no body parameter

7.5. draft-ietf-eai-rfc5336bis: Version 03

improve the text

7.6. draft-ietf-eai-rfc5336bis: Version 04

update the abstract

improve the text

7.7. draft-ietf-eai-rfc5336bis: Version 05

improve the text based on AD and Co-chairs

7.8. draft-ietf-eai-rfc5336bis: Version 06

update the iana consideration

7.9. draft-ietf-eai-rfc5336bis: Version 07

improve the iana consideration

7.10. draft-ietf-eai-rfc5336bis: Version 08

improve the texts

add the mail parameter

add the new section about mail command parameter usage

update the security consideration

7.11. draft-ietf-eai-rfc5336bis: Version 09

improve the texts

7.12. draft-ietf-eai-rfc5336bis: Version 10

refine the ABNF definitions

improve the texts

7.13. draft-ietf-eai-rfc5336bis: Version 11

remove the update of RFC5321 and RFC5322

change the title from "SMTP Extension for Internationalized Email Address" to "SMTP Extension for Internationalized Email" based on Ernie's comment

the trace field of section 3.7.3 is updated to reflect the WG's conclusion

improve the texts

8. References

8.1. Normative References

- [ASCII] American National Standards Institute (formerly United States of America Standards Institute), "USA Code for Information Interchange", ANSI X3.4-1968, 1968.
- [RFC2033] Myers, J., "Local Mail Transfer Protocol", RFC 2033, October 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3463] Vaudreuil, G., "Enhanced Mail System Status Codes", RFC 3463, January 2003.
- [RFC3464] Moore, K. and G. Vaudreuil, "An Extensible Message Format for Delivery Status Notifications", RFC 3464, January 2003.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", RFC 3629, November 2003.

- [RFC4409] Gellens, R. and J. Klensin, "Message Submission for Mail", RFC 4409, April 2006.
- [RFC4952bis]
Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", I-D rfc4952bis, September 2010.
- [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.
- [RFC5248] Hansen, T. and J. Klensin, "A Registry for SMTP Enhanced Mail System Status Codes", RFC 5248, June 2008.
- [RFC5321] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, October 2008.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.
- [RFC5335bis]
Abel, Y. and S. Steel, "Internationalized Email Headers", I-D rfc5335bis, March 2011.
- [RFC5337bis]
Hansen, T., Ed., Newman, C., and A. Melnikov, Ed., "Internationalized Delivery Status and Disposition Notifications", I-D 5337bis, October 2010.
- [RFC5890] Klensin, J., "Internationalizing Domain Names in Applications (IDNA definitions)", RFC 5890, June 2010.
- [RFC6152] Klensin, J., Freed, N., Rose, M., and D. Crocker, "SMTP Service Extension for 8-bit MIME Transport", STD 71, RFC 6152, March 2011.

8.2. Informative References

- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC3030] Vaudreuil, G., "SMTP Service Extensions for Transmission of Large and Binary MIME Messages", RFC 3030, December 2000.

- [RFC3207] Hoffman, P., "SMTP Service Extension for Secure SMTP over Transport Layer Security", RFC 3207, February 2002.
- [RFC4954] Siemborski, R. and A. Melnikov, "SMTP Service Extension for Authentication", RFC 4954, July 2007.
- [RFC5336] Yao, J. and W. Mao, "SMTP Extension for Internationalized Email Addresses", RFC 5336, September 2008.
- [RFC5598] Crocker, D., "Internet Mail Architecture", RFC 5598, July 2009.

Authors' Addresses

Jiankang YAO
CNNIC
No.4 South 4th Street, Zhongguancun
Beijing

Phone: +86 10 58813007
Email: yaojk@cnnic.cn

Wei MAO
CNNIC
No.4 South 4th Street, Zhongguancun
Beijing

Phone: +86 10 58812230
Email: maowei_ietf@cnnic.cn

Network Working Group
Internet-Draft
Obsoletes: RFC5721 (if approved)
Intended status: Standards Track
Expires: January 12, 2012

R. Gellens
QUALCOMM Incorporated
C. Newman
Oracle
J. Yao
CNNIC
K. Fujiwara
JPRS
July 11, 2011

POP3 Support for UTF-8
draft-ietf-eai-rfc5721bis-02.txt

Abstract

This specification extends the Post Office Protocol version 3 (POP3) to support un-encoded international characters in user names, passwords, mail addresses, message headers, and protocol-level textual strings.

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 January 12, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	3
1.1. Conventions Used in This Document	3
2. LANG Capability	4
3. UTF8 Capability	6
3.1. The UTF8 Command	7
3.2. USER Argument to UTF8 Capability	8
4. Native UTF-8 Maildrops	9
5. IANA Considerations	9
6. Security Considerations	9
7. Change History	10
7.1. draft-ietf-eai-rfc5721bis: Version 00	10
7.2. draft-ietf-eai-rfc5721bis: Version 01	10
7.3. draft-ietf-eai-rfc5721bis: Version 02	10
8. References	10
8.1. Normative References	10
8.2. Informative References	12
Appendix A. Design Rationale	12
Appendix B. Acknowledgments	12

1. Introduction

This document forms part of the Email Address Internationalization (EAI) protocols described in the EAI Framework document [I-D.ietf-eai-frmwrk-4952bis]. As part of the overall EAI work, email messages may be transmitted and delivered containing un-encoded UTF-8 characters, and mail drops that are accessed using POP3 [RFC1939] might natively store UTF-8.

This specification extends POP3 [RFC1939] using the POP3 extension mechanism [RFC2449] to permit un-encoded UTF-8 [RFC3629] in headers, as described in "Internationalized Email Headers" [I-D.ietf-eai-rfc5335bis]. It also adds a mechanism to support login names and passwords with UTF-8 charset beyond ASCII, and a mechanism to support UTF-8 charset in protocol level response strings for languages beyond English.

Within this specification, the term "down-conversion" refers to the process of modifying a message containing UTF-8 headers [I-D.ietf-eai-rfc5335bis] or body parts with 8bit content-transfer-encoding, as defined in MIME Section 2.8 [RFC2045], into conforming 7-bit Internet Message Format [RFC5322] with message header extensions for non-ASCII text [RFC2047] and other 7-bit encodings. Down-conversion is specified by "Post-delivery Message Downgrading for Internationalized Email Messages" [popimap-downgrade].

1.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

In examples, "C:" and "S:" indicate lines sent by the client and server, respectively. If a single "C:" or "S:" label applies to multiple lines, then the line breaks between those lines are for editorial clarity only and are not part of the actual protocol exchange.

Note that examples always use 7-bit ASCII characters due to limitations of this document format; in particular, some examples for the "LANG" command may appear silly as a result.

2. LANG Capability

Per "POP3 Extension Mechanism" [RFC2449], this document adds a new capability response tag to indicate support for a new command: LANG. The capability tag and new command are described below.

CAPA tag:
LANG

Arguments with CAPA tag:
none

Added Commands:
LANG

Standard commands affected:
All

Announced states / possible differences:
both / no

Commands valid in states:
AUTHORIZATION, TRANSACTION

Specification reference:
this document

Discussion:

POP3 allows most +OK and -ERR server responses to include human-readable text that, in some cases, might be presented to the user. But that text is limited to ASCII by the POP3 specification [RFC1939]. The LANG capability and command permit a POP3 client to negotiate which language the server should use when sending human-readable text.

A server that advertises the LANG extension MUST use the language "i-default" as described in [RFC2277] as its default language until another supported language is negotiated by the client. A server MUST include "i-default" as one of its supported languages.

The LANG command requests that human-readable text included in all subsequent +OK and -ERR responses be localized to a language matching the language range argument (the "Basic Language Range" as described by [RFC4647]). If the command succeeds, the server returns a +OK response followed by a single space, the exact language tag selected, another space, and the rest of the line is human-readable text in the appropriate language. This and subsequent protocol-level human-

readable text is encoded in the UTF-8 charset.

If the command fails, the server returns an -ERR response and subsequent human-readable response text continues to use the language that was previously active (typically i-default).

The special "*" language range argument indicates a request to use a language designated as preferred by the server administrator. The preferred language MAY vary based on the currently active user.

If no argument is given and the POP3 server issues a positive response, then the response given is multi-line. After the initial +OK, for each language tag the server supports, the POP3 server responds with a line for that language. This line is called a "language listing".

In order to simplify parsing, all POP3 servers are required to use a certain format for language listings. A language listing consists of the language tag [RFC5646] of the message, optionally followed by a single space and a human-readable description of the language in the language itself, using the UTF-8 charset.

Examples:

< Note that some examples do not include the correct character accents due to limitations of this document format. >

< The server defaults to using English i-default responses until the client explicitly changes the language. >

```
C: USER karen
S: +OK Hello, karen
C: PASS password
S: +OK karen's maildrop contains 2 messages (320 octets)
```

< Client requests deprecated MUL language. Server replies with -ERR response. >

```
C: LANG MUL
S: -ERR invalid language MUL
```

< A LANG command with no parameters is a request for a language listing. >

```
C: LANG
S: +OK Language listing follows:
S: en English
S: en-boont English Boontling dialect
```

```
S: de Deutsch
S: it Italiano
S: es Espanol
S: sv Svenska
S: i-default Default language
S: .
```

< A request for a language listing might fail. >

```
C: LANG
S: -ERR Server is unable to list languages
```

< Once the client changes the language, all responses will be in that language, starting with the response to the LANG command. >

```
C: LANG es
S: +OK es Idioma cambiado
```

< If a server does not support the requested primary language, responses will continue to be returned in the current language the server is using. >

```
C: LANG uga
S: -ERR es Idioma <<UGA>> no es conocido
```

```
C: LANG sv
S: +OK sv Kommandot "LANG" lyckades
```

```
C: LANG *
S: +OK es Idioma cambiado
```

3. UTF8 Capability

Per "POP3 Extension Mechanism" [RFC2449], this document adds a new capability response tag to indicate support for new server functionality, including a new command: UTF8. The capability tag and new command and functionality are described below.

CAPA tag:
UTF8

Arguments with CAPA tag:
USER

Added Commands:
UTF8

Standard commands affected:

USER, PASS, APOP, LIST, TOP, RETR

Announced states / possible differences:

both / no

Commands valid in states:

AUTHORIZATION

Specification reference:

this document

Discussion:

This capability adds the "UTF8" command to POP3. The UTF8 command switches the session from ASCII to UTF-8 mode. In UTF-8 mode, it means that both servers and clients can send and accept the UTF-8 characters.

3.1. The UTF8 Command

The UTF8 command enables UTF-8 mode. The UTF8 command has no parameters.

Maildrops can natively store UTF-8 or be limited to ASCII. UTF-8 mode has no effect on messages in an ASCII-only maildrop. Messages in native UTF-8 maildrops can be ASCII or UTF-8 using internationalized headers [I-D.ietf-eai-rfc5335bis] and/or 8bit content-transfer-encoding, as defined in MIME Section 2.8 [RFC2045]. In UTF-8 mode, both UTF-8 and ASCII messages are sent to the client as-is (without conversion). When not in UTF-8 mode, UTF-8 messages in a native UTF-8 maildrop MUST NOT be sent to the client as-is. The UTF8 Commands MAY fail. UTF-8 messages in a native UTF-8 maildrop MAY be down-converted (downgraded) to comply with unextended POP and Internet Mail Format without UTF-8 mode support.

Note that even in UTF-8 mode, MIME binary content-transfer-encoding is still not permitted.

The octet count (size) of a message reported in a response to the LIST command SHOULD match the actual number of octets sent in a RETR response (not counting byte-stuffing). Sizes reported elsewhere, such as in STAT responses and non-standardized, free-form text in positive status indicators (following "+OK") need not be accurate, but it is preferable if they were.

Mail stores are either ASCII or native UTF-8, and clients either issue the UTF8 command or not. The message needs converting only

when it is native UTF-8 and the client has not issued the UTF8 command, in which case the server MAY choose to down-convert it. The down-converted message may be larger. The server may choose various strategies regarding down-conversion, which include when to down-convert, whether to cache or store the down-converted form of a message (and if so, for how long), and whether to calculate or retain the size of a down-converted message independently of the down-converted content. If the server does not have immediate access to the accurate down-converted size, it may be faster to estimate rather than calculate it. Servers are expected to normally follow the RFC 1939 [RFC1939] text on using the "exact size" in a scan listing, but there may be situations with maildrops containing very large numbers of messages in which this might be a problem. If the server does estimate, reporting a scan listing size smaller than what it turns out to be could be a problem for some clients. In summary, it is better for servers to report accurate sizes, but if this is not possible, high guesses are better than small ones. Some POP servers include the message size in the non-standardized text response following "+OK" (the 'text' production of RFC 2449 [RFC2449]), in a RETR or TOP response (possibly because some examples in POP3 [RFC1939] do so). There has been at least one known case of a client relying on this to know when it had received all of the message rather than following the POP3 [RFC1939] rule of looking for a line consisting of a termination octet (".") and a CRLF pair. While any such client is non-compliant, if a server does include the size in such text, it is better if it is accurate.

Clients MUST NOT issue the STLS command [RFC2595] after issuing UTF8; servers MAY (but are not required to) enforce this by rejecting with an "-ERR" response an STLS command issued subsequent to a successful UTF8 command. (Because this is a protocol error as opposed to a failure based on conditions, an extended response code [RFC2449] is not specified.)

3.2. USER Argument to UTF8 Capability

If the USER argument is included with this capability, it indicates that the server accepts UTF-8 user names and passwords.

Servers that include the USER argument in the UTF8 capability response SHOULD apply SASLprep [RFC4013] to the arguments of the USER and PASS commands.

A client or server that supports APOP and permits UTF-8 in user names or passwords MUST apply SASLprep [RFC4013] to the user name and password used to compute the APOP digest.

When applying SASLprep [RFC4013], servers MUST reject UTF-8 user

names or passwords that contain a Unicode character listed in Section 2.3 of SASLprep [RFC4013]. When applying SASLprep to the USER argument, the PASS argument, or the APOP username argument, a compliant server or client MUST treat them as a query string [RFC3454] (i.e., unassigned Unicode code points are allowed). When applying SASLprep to the APOP password argument, a compliant server or client MUST treat them as a stored string [RFC3454] (i.e., unassigned Unicode code points are prohibited).

The client does not need to issue the UTF8 command prior to using UTF-8 in authentication. However, clients MUST NOT use UTF-8 characters in USER, PASS, or APOP commands unless the USER argument is included in the UTF8 capability response.

The server MUST reject UTF-8 user names or passwords that fail to comply with the formal syntax in UTF-8 [RFC3629].

Use of UTF-8 characters in the AUTH command is governed by the POP3 SASL [RFC5034] mechanism.

4. Native UTF-8 Maildrops

When a POP3 server uses a native UTF-8 maildrop, it is the responsibility of the server to comply with the POP3 base specification [RFC1939] and Internet Message Format [RFC5322] when not in UTF-8 mode. Mechanisms for 7-bit downgrading to help comply with the standards are described in [popimap-downgrade].

5. IANA Considerations

This specification updates two capabilities ("UTF8" and "LANG") to the POP3 capability registry [RFC2449].

6. Security Considerations

The security considerations of UTF-8 [RFC3629] and SASLprep [RFC4013] apply to this specification, particularly with respect to use of UTF-8 in user names and passwords.

The "LANG *" command might reveal the existence and preferred language of a user to an active attacker probing the system if the active language changes in response to the USER, PASS, or APOP commands prior to validating the user's credentials. Servers MUST implement a configuration to prevent this exposure.

It is possible for a man-in-the-middle attacker to insert a LANG command in the command stream, thus making protocol-level diagnostic responses unintelligible to the user. A mechanism to integrity-

protect the session, such as Transport Layer Security (TLS) [RFC2595] can be used to defeat such attacks.

Modifying server authentication code (in this case, to support UTF8 command) needs to be done with care to avoid introducing vulnerabilities (for example, in string parsing).

The UTF8 command description (Section 3.1) contains a discussion on reporting inaccurate sizes. An additional risk to doing so is that, if a client allocates buffers based on the reported size, it may overrun the buffer, crash, or have other problems if the message data is larger than reported.

7. Change History

7.1. draft-ietf-eai-rfc5721bis: Version 00

following the new charter

7.2. draft-ietf-eai-rfc5721bis: Version 01

refine the texts

7.3. draft-ietf-eai-rfc5721bis: Version 02

update the texts based on Joseph's comments

8. References

8.1. Normative References

- [I-D.ietf-eai-frmwrk-4952bis] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", draft-ietf-eai-frmwrk-4952bis-10 (work in progress), September 2010.
- [I-D.ietf-eai-rfc5335bis] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", draft-ietf-eai-rfc5335bis-11 (work in progress), July 2011.
- [RFC1939] Myers, J. and M. Rose, "Post Office Protocol - Version 3", STD 53, RFC 1939, May 1996.
- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions

- (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.
- [RFC2449] Gellens, R., Newman, C., and L. Lundblade, "POP3 Extension Mechanism", RFC 2449, November 1998.
- [RFC3454] Hoffman, P. and M. Blanchet, "Preparation of Internationalized Strings ("stringprep")", RFC 3454, December 2002.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.
- [RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", RFC 4013, February 2005.
- [RFC4647] Phillips, A. and M. Davis, "Matching of Language Tags", BCP 47, RFC 4647, September 2006.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.
- [RFC5646] Phillips, A. and M. Davis, "Tags for Identifying Languages", BCP 47, RFC 5646, September 2009.

8.2. Informative References

- [RFC2595] Newman, C., "Using TLS with IMAP, POP3 and ACAP", RFC 2595, June 1999.
- [RFC5034] Siemborski, R. and A. Menon-Sen, "The Post Office Protocol (POP3) Simple Authentication and Security Layer (SASL) Authentication Mechanism", RFC 5034, July 2007.
- [popimap-downgrade] Fujiwara, K., "Post-delivery Message Downgrading for Internationalized Email Messages", draft-ietf-eai-popimap-downgrade-00 (work in progress), October 2010.

Appendix A. Design Rationale

This non-normative section discusses the reasons behind some of the design choices in the above specification.

Due to interoperability problems with RFC 2047 and limited deployment of RFC 2231, it is hoped these 7-bit encoding mechanisms can be deprecated in the future when UTF-8 header support becomes prevalent.

USER is optional because the implementation burden of SASLprep [RFC4013] is not well understood, and mandating such support in all cases could negatively impact deployment.

Appendix B. Acknowledgments

Thanks to John Klensin, Tony Hansen, and other EAI working group participants who provided helpful suggestions and interesting debate that improved this specification.

Authors' Addresses

Randall Gellens
QUALCOMM Incorporated
5775 Morehouse Drive
San Diego, CA 92651
US

EMail: rg+ietf@qualcomm.com

Chris Newman
Oracle
800 Royal Oaks
Monrovia, CA 91016-6347
US

EMail: chris.newman@oracle.com

Jiankang YAO
CNNIC
No.4 South 4th Street, Zhongguancun
Beijing

Phone: +86 10 58813007
EMail: yaojk@cnnic.cn

Kazunori Fujiwara
Japan Registry Services Co., Ltd.
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Tokyo

Phone: +81 3 5215 8451
EMail: fujiwara@jprs.co.jp

