

Email Address Internationalization
(EAI)
Internet-Draft
Intended status: Standards Track
Expires: October 20, 2011

K. Fujiwara
JPRS
Apr 18, 2011

Post-delivery Message Downgrading for Internationalized Email Messages
draft-ietf-eai-popimap-downgrade-01.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/1id-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 October 20, 2011.

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.	New Header Fields Definition	4
3.1.	Unknown Header Fields' Preservation Header Fields	5
4.	Email Header Fields Downgrading	5
4.1.	Downgrading Method for Each ABNF Element	5
4.1.1.	RECEIVED Downgrading	6
4.1.2.	UNSTRUCTURED Downgrading	6
4.1.3.	WORD Downgrading	6
4.1.4.	COMMENT Downgrading	6
4.1.5.	MIME-VALUE Downgrading	6
4.1.6.	DISPLAY-NAME Downgrading	6
4.1.7.	MAILBOX Downgrading	6
4.1.8.	ENCAPSULATION Downgrading	7
4.1.9.	TYPED-ADDRESS Downgrading	7
4.2.	Downgrading Method for Each Header Field	7
4.2.1.	Address Header Fields That Contain <address>s	7
4.2.2.	Address Header Fields with Typed Addresses	8
4.2.3.	Downgrading Non-ASCII in Comments	8
4.2.4.	Received Header Field	9
4.2.5.	MIME Content Header Fields	9
4.2.6.	Non-ASCII in <unstructured>	9
4.2.7.	Non-ASCII in <phrase>	9
4.2.8.	Other Header Fields	9
5.	MIME Body-Part Header Field Downgrading	10
6.	Security Considerations	10
7.	Implementation Notes	11
7.1.	RFC 2047 Encoding	11
8.	IANA Considerations	11
9.	Acknowledgements	11
10.	References	12
10.1.	Normative References	12
10.2.	Informative References	13
Appendix A.	Examples	13
A.1.	Downgrading Example	13

Fujiwara

Expires October 20, 2011

[Page 2]

1. Introduction

Traditional mail systems, which are defined by [\[RFC5322\]](#), allow ASCII characters in mail header field values. The UTF8SMTP extension ([\[I-D.ietf-eai-frmwrk-4952bis\]](#) 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 converts mail header fields and MIME header fields to ASCII.

This document does not change any protocols except by defining new header fields. It describes the conversion method from the internationalized email messages that are defined in [\[I-D.ietf-eai-frmwrk-4952bis\]](#), and [\[I-D.ietf-eai-rfc5335bis\]](#) to the traditional email messages defined in [\[RFC5322\]](#).

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 three parts:

- o New header field definitions
- o Email header field downgrading
- o MIME header field downgrading

In [Section 3](#) of this document, header fields starting with "Downgraded-" are introduced. They preserve the original header fields.

Email header field downgrading is described in [Section 4](#). 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 5](#). 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 [\[I-D.ietf-eai-frmwrk-4952bis\]](#), 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", "i18mail address", "UTF8SMTP", "message", and "mailing list" are used with the definitions from [\[I-D.ietf-eai-frmwrk-4952bis\]](#).

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. 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.

3.1. 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.

4. 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>.

4.1. 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.

4.1.1. 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.

4.1.2. UNSTRUCTURED Downgrading

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

4.1.3. WORD Downgrading

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

4.1.4. COMMENT Downgrading

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

4.1.5. 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.

4.1.6. 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.

4.1.7. 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](#)].

[[Note: If the original non-ASCII address is a part of a group address, this rewriting may conflict the original DISPLAY-NAME. This problem need to be fixed.]]

4.1.8. 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 3.1](#) as a last resort.

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

4.1.9. 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.

4.2. 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.

4.2.1. 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 <mailbox> elements that contain non-ASCII addresses, perform COMMENT downgrading, DISPLAY-NAME downgrading, and MAILBOX downgrading.

[[Note: [RFC 5322](#) does not allow group syntax in "From:", "Resent-From:", "Sender:", "Resent-Sender:", but proposed method uses group syntax. This problem need to be fixed.]]

4.2.2. Address Header Fields with Typed Addresses

Original-Recipient:
Final-Recipient:

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

4.2.3. 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.

4.2.4. Received Header Field

Received:

Perform COMMENT downgrading and RECEIVED downgrading.

4.2.5. MIME Content Header Fields

Content-Type:
Content-Disposition:

Perform MIME-VALUE downgrading and COMMENT downgrading.

4.2.6. Non-ASCII in <unstructured>

Subject:
Comments:
Content-Description:

Perform UNSTRUCTURED downgrading.

4.2.7. Non-ASCII in <phrase>

Keywords:

Perform WORD downgrading.

4.2.8. 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.

5. 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 4](#).

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.

6. 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 [\[I-D.ietf-eai-frmwrk-4952bis\]](#) for more discussion.

7. Implementation Notes

7.1. [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.

8. 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 3.1](#) to avoid conflicts with existing IETF activity (Email Address Internationalization).

9. 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.

10. References

10.1. 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. and S. Steele, "Internationalized Email Headers", [draft-ietf-eai-rfc5335bis-10](#) (work in progress), March 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.
- [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.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", [RFC 5322](#), October 2008.
- [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.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

[10.2.](#) 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