Email Address Internationalization (EAI)
Internet-Draft
Expires: November 27, 2006

Y. YONEYA, Ed.
K. Fujiwara, Ed.
JPRS
May 26, 2006

Downgrading mechanism for Internationalized eMail Address (IMA)
draft-ietf-eai-downgrade-00.txt

Status of this Memo

Copyright Notice

Abstract

Traditional mail system handles only US-ASCII characters in SMTP envelope and mail headers. The Internationalized eMail Address (IMA) is implemented by allowing UTF-8 characters in SMTP envelope and mail headers. To deliver IMA through IMA incompliant environment, some sort of converting mechanism (i.e. downgrading) is required. This document describes requirements for downgrading, SMTP session downgrading, header downgrading and implementation consideration.

Table of Contents

1. Introduction

Traditional mail system which is defined by [RFC2821] and [RFC2822] allows US-ASCII characters in SMTP envelop and mail headers. IMA proposal [IMA-overview],[IMA-UTF8], [IMA-SMTPext] allows UTF-8 characters in SMTP envelop and mail headers.

Carrying IMA from sender to recipients requires all components on the mail delivery route are IMA compliant. Otherwise IMA can't be delivered. To solve the problem, this document describes downgrading mechanism that enables delivering IMA by converting it to corresponding US-ASCII representation on current mail delivery system. Not only SMTP envelope, but also UTF-8 in mail headers MUST be converted to US-ASCII.

Downgrading in IMA consists from following two parts:
- SMTP session downgrade
- header downgrade

Decoding downgraded envelope/message is called 'Upgrading' in this document. Each downgrading mechanism has corresponding upgrading mechanism.

In this document, requirements for downgrading is described in section Section 3, SMTP session downgrade is described in Section 4, and header downgrade is described in Section 5.

2. Terminology

This document assumes a reasonable understanding of the protocols and terminology of the core email standards as documented in [RFC2821] and [RFC2822].

Much of the description in this document depends on the abstractions of "Mail Transfer Agent" ("MTA") and "Mail User Agent" ("MUA"). However, it is important to understand that those terms and the underlying concepts postdate the design of the Internet's email architecture and the "protocols on the wire" principle. That email architecture, as it has evolved, and the "wire" principle have prevented any strong and standardized distinctions about how MTAs and MUAs interact on a given origin or destination host (or even whether they are separate).

The final ("delivery") MTA stores Mail messages in a "message store" or resends messages where the receiver has assigned. In this document, this function is called Mail Delivery Agent(MDA).

In this document, an address is "all-ASCII" if every character in the address is in the ASCII character repertoire [ASCII]; an address is "non-ASCII" if any character is not in the ASCII character repertoire. The term "all-ASCII" is also applied to other protocol elements when the distinction is important, with "non-ASCII" or "internationalized" as its opposite.

The key words "MUST", "SHALL", "REQUIRED", "SHOULD", "RECOMMENDED", and "MAY" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Downgrade Requirements

3.1. Timing and conditions of downgrading

This section describes timing and conditions of downgrading.

- Timing: SMTP client detects SMTP server doesn't support "IEmail" option at EHLO. [IMA-SMTPext]
- Conditions: SMTP client detects that UTF-8 is included in SMTP envelope or mail headers.

Note: If the i18n-email header exists, downgrading will be performed. If UTF-8 headers are present without the i18n-email header, this is a protocol error, and handling of this situation is outside the scope of this specification.

3.2. Requirements

1. Downgrading must be performed only once.
2. Upgrading must be performed at minimized place such as final destination like recipient MUA.
3. Downgrading and decoding must be automated.
4. Downgrading and decoding should be easy and lightweight as it is possible to do with MTA like 8BITMIME encapsulation.
5. Downgrade and upgrade method must be defined clearly.
6. Downgrading and decoding should preserve all header information.
7. Downgrading must support SPF and DKIM.
8. Downgrading occurrence must be recorded.

4. SMTP Downgrading

Downgrading MUST be performed in each SMTP session. Target of downgrading elements in SMTP envelope are below:

- MAIL FROM:
- RCPT TO:

Downgrading in SMTP envelope uses ALT-ADDR and ATOMIC option proposed in [IMA-SMTPext].

If downgrading is expected, mail sender MUA MUST append ALT-ADDR or ATOMIC option to all IMA envelope addresses to denote alternative US-ASCII address when sending mail.

When MUA/MTA is transferring mail and finding its envelope is IMA, it MUST decide to bounce or downgrade if receiving MTA is IMA incompliant.

Both ALT-ADDR parameter and ATOMIC parameter is specified in one envelope from/to, use ALT-ADDR parameter and ignore ATOMIC parameter.

Further, even if no downgrading is performed for envelope from/to, MUA/MTA MUST downgrade headers including UTF-8 or bounce. This is described in next section.

MTA MAY downgrade messages that envelope from/to of IMA have ALT-ADDR with alternative US-ASCII address or ATOMIC is "y".

MTA generates alternative US-ASCII address when ALT-ADDR option is not specified and ATOMIC is "y".

Alternative US-ASCII address generation algorithms are below:
domain-part: Punycode/IDNA [RFC3490]
local-part: Punycode[RFC3492] without normalization. Prefix MUST be assigned by IANA (which is not "xn--").

MTA replaces IMA with specified or generated alternative US-ASCII address. Then appends replaced information with IMA-Downgraded-From and IMA-Downgraded-To header in mail header (outgoing SMTP DATA).

```
IMA-Downgraded-From: <IMA> <US-ASCII>
IMA-Downgraded-To: <IMA> <US-ASCII>
```

Note that when downgrading, not to disclose whole recipient address, MUA/MTA SHOULD make SMTP connection per each recipient address.

Also note that by appending IMA-Downgraded-From/To headers, MUA/MTA MUST perform SMTP/Header downgrading. This is described in next section.

Downgraded envelope to is parsed only in MDA and delivered to final mailbox.

Case study: SPF check

SPF checks envelope from's domainname and smtp connection IP address. If ALT-ADDR's domainname is Punycode/IDNA of IMA domainname, it is equal to SPF/IMA (need to define). In this case, SPF check will be performed correctly. Otherwise, more detailed consideration is required.

5. SMTP DATA/Header downgrading

In this section, four methods for SMTP DATA/Header downgrading is proposed. Working group should select one.

- No header downgrading
- Encapsulating whole SMTP DATA
- Translating each header
- Encapsulating each header

Target and non-target of downgrading elements in mail headers (SMTP data) are below:

Originator address(es): IMA in From, Reply-To, Sender and their Resent- headers MUST be target of downgrading.
Destination address(es): IMA in To, CC, Bcc and their Resent- headers MUST be target of downgrading.
IDs: IDs such as Message-ID, Date, In-Reply-To and References MUST NOT be target of downgrading.
Trace headers: Received headers which contains IMA MUST be target of downgrading.
other headers: UTF-8 in other headers MUST be target of downgrading.

Rewriting Received header is prohibited in [RFC2821] Section 4.4 Trace field. But downgrading may be considered as the 'Mail Gatewaying' which is described in [RFC2821] Section 3.8. If it is true, these downgrading methods are acceptable.

5.1. No header downgrading

Most MTAs support 8bit characters in mail headers. Currently, mail systems in some countries or languages use raw 8bit header value in their local encoding. This method does not care about using UTF-8 headers in existing mail systems.

This method may break existing mail infrastructure.

5.2. Downgrading with MIME encapsulation

This downgrade method requires new MIME 'Content-Type:' which express EAI(Email Address Internationalization). This document assumes 'Content-Type: Message/EAI' existence.

Downgrading:

* If mail header contains UTF-8 data, downgrade whole message to be MIME encoded. Whole message becomes new MIME part (Message/EAI).
* Originator Addresses (From, Sender, etc.), Destination Addresses (To, CC, etc.), IDs (Message-ID, etc.), Subject, Date headers are copied from original header.
* If From header contains IMA, it is replaced with downgraded Envelope-from.
* If To or CC headers contain IMA, they are replaced with single downgraded envelope-to as To header.
* If Subject header contains UTF-8, it is replaced to a certain message or encoded by RFC2047.
* Message-ID, Date headers are preserved.

As a result, new body contains one new MIME part (Message/EAI).

Encoding example

```
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Content-Type: Multipart/Mixed;
   boundary="--Next_Part(unique_string)--"
Content-Transfer-Encoding: 8bit
Subject: DOWNGRADED_SUBJECT
From: DOWNGRADED_FROM
To: DOWNGRADED_TO
Date: DATE

----Next_Part(unique_string)--
Content-Type: Message/EAI
Content-Transfer-Encoding: 8bit
Content-Disposition: inline

IMA-Downgraded-From: <IMA> <DOWNGRADED_FROM>
IMA-Downgraded-To: <IMA> <DOWNGRADED_TO>
Received: ... for IMA
Received: ... for IMA
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Subject: UTF-8_SUBJECT
From: IMA
To: IMA
Date: DATE

MAIL_BODY

----Next_Part(unique_string)----
```

Figure 1

Upgrading:

* If mail message contains only one MIME part and its Content-Type is 'Message/EAI', it may be a downgraded message. To check if downgraded, compare mail body's message-id and MIME part's message-id. If message-ids are the same, it is downgraded message. Then, treat MIME part as entire mail message.
* When checking trace field, checker SHOULD check Received header both in wrapping headers and headers in encapsulated part.

Case study: DKIM

DKIM checker performs decoding downgraded message first.

Pros:

* MTA does not need to decode each headers carefully.
* Whole headers can be submitted AS IS.

Cons:

* IMA from/to can not distinguish from encoded mail headers.
* IMA incompliant MUA can not treat encoded message.

5.3. Header conversion

Define conversion method to US-ASCII for all headers which contains IMA.

Each header has its own downgrading method. Basically, MIME encoding of RFC 2047. Recipient/Sender addresses and Received headers which may contain IMA need special processing.

Downgrading:

From, To, CC, Resent-From, Resent-To headers which contains Originator/Destination address(es): Extract every addr-spec [RFC2822] of mailboxes which includes UTF-8 characters. For each addr-spec, if it includes UTF-8, convert it into ACE with the same method described in Section 4. Original IMA SHOULD remain as a comment encoded by MIME with UTF-8 tag [RFC2047]. Note that some special characters in addr-spec MUST be escaped. If mailbox elements except for addr-spec include UTF-8, those MUST be encoded by base64 with UTF-8 tag.

Downgrading other header: Encode UTF-8 characters of headers by MIME with UTF-8 tag [RFC2047].

Upgrading: If each mail header has [RFC2047] encoded part and which encoding is "UTF-8", it is downgraded header.

Pros:

* IMA incompliant MUA can display mail body except for original IMA from/to.

Cons:

* Implementation is difficult because MUA/MTA must parse each header and encode it by defined method.
* Hard to preserve whole information AS IS. Therefore, to check DKIM requires special consideration.

5.4. Translating each header

Define generic encapsulation header: "Downgraded: HeaderName: HeaderValue". Header value is encoded in [RFC2047] with UTF-8 tag.

Downgrading:
All headers which contains UTF-8 characters are encapsulated to generic encapsulation header. There is no special handling for recipient/sender addresses in From, To, CC, Resent-* headers. Received headers need special consideration.
If downgrading process encapsulates From header, downgrade process should generate From header from the envelope from address with downgraded mark in comment field.
If downgrading process encapsulates all To, CC headers, downgrade process should generate To header from the envelope to address with downgraded mark in comment field.

Upgrading:
If each mail header has [RFC2047] encoded part and which encoding is "UTF-8", it is downgraded header and the upgrading process decode this header.

Pros:
* IMA incompliant MUA can display mail body except for original IMA from/to.
* Implementation is easier than Section 5.3

Cons:
* This method may break [RFC2821] [RFC2821].
* Hard to preserve whole information AS IS. Therefore, to check DKIM requires special consideration.

6. Implementation consideration

6.1. MUA

IMA compliant MUA MUST implement downgrade mechanism for sending.

MUA MAY encode UTF-8 in Subject header with the same encoding of body part while downgrading.

IMA compliant MUA MUST decode downgraded mail and MUST show IMA on display.

6.2. MDA Requirements

This section describes downgrading in MDA.
1. MDA MUST NOT convert downgraded header to UTF-8.
2. Record Return-Path header in ACE form.
3. Perform downgrading for each Storage/Back-end-Process. If and only if MDA knows MUA is IMA compliant, then no downgrading is performed.

4. If MDA detects that SMTP recipient address is downgraded IMA, then MDA MUST decode IMA and perform the same processing as if it were IMA. MDA MAY normalize or canonicalize local-part before processing it.

7. Security considerations

See the extended security considerations discussion in [IMA-overview]

8. IANA Considerations

To distinguish downgraded IMA in ACE form, it MUST have ACE-Prefix. The ACE-Prefix MUST differ from IDNA ACE-Prefix to avoid possible confusion. IANA will assign IMA ACE-Prefix when RFC is published.

9. Acknowledgements

John Klensin, Harald Alvestrand, Chris Newman, Charles Lindsey, Marcos Sanz, Alexey Melnikov, and JET members.

10. Normative References

[ASCII] American National Standards Institute (formerly United States of America Standards Institute), "USA Code for Information Interchange", ANSI X3.4-1968, 1968.

ANSI X3.4-1968 has been replaced by newer versions with slight modifications, but the 1968 version remains definitive for the Internet.

[Hoffman-IMAA] Hoffman, P. and A. Costello, "Internationalizing Mail Addresses in Applications (IMAA)", draft-hoffman-imaa-03 (work in progress), October 2003.

[IMA-Constraints] Klensin, J., "Internationalization in Internet Applications: Issues, Tradeoffs, and Email Addresses", draft-klensin-ima-constraints-00 (work in progress), Febrary 2006.

[IMA-SMTPext] Yao, J., Ed., "SMTP extension for internationalized email address", draft-ietf-eai-smtpext-00 (work in progress), Febrary 2006.

[IMA-UTF8] Yeh, J., "Internationalized Email Headers", draft-yeh-ima-utf8headers-01 (work in progress), February 2006.

[IMA-overview] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", draft-ietf-eai-framework-00 (work in progress), May 2006.

[JET-IMA] Yao, J. and J. Yeh, "Internationalized eMail Address (IMA)", draft-lee-jet-ima-00 (work in progress), June 2005.

[Klensin-emailaddr] Klensin, J., "Internationalization of Email Addresses", draft-klensin-emailaddr-i18n-03 (work in progress), July 2005.

[RFC1123] Braden, R., "Requirements for Internet Hosts - Application and Support", STD 3, RFC 1123, October 1989.

[RFC1651] Klensin, J., Freed, N., Rose, M., Stefferud, E., and D. Crocker, "SMTP Service Extensions", RFC 1651, July 1994.

[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.

[RFC2821] Klensin, J., "Simple Mail Transfer Protocol", RFC 2821, April 2001.

[RFC2822] Resnick, P., "Internet Message Format", RFC 2822, April 2001.

[RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.

[RFC3492] Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.

[RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.

Authors' Addresses

Yoshiro YONEYA (editor)
JPRS
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Chiyoda-ku, Tokyo 101-0065
Japan

Phone: +81 3 5215 8451
Email: yone@jprs.co.jp

Kazunori Fujiwara (editor)
JPRS
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

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

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