Downgrading mechanism for Email Address Internationalization (EAI)
draft-ietf-eai-downgrade-01.txt

Status of this Memo

   By submitting this Internet-Draft, each author represents that any
   applicable patent or other IPR claims of which he or she is aware
   have been or will be disclosed, and any of which he or she becomes
   aware will be disclosed, in accordance with Section 6 of 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 December 28, 2006.

Copyright Notice

Abstract

   Traditional mail systems handle only US-ASCII characters in SMTP
   envelope and mail headers. The Email Address Internationalization
   (EAI) is implemented by allowing UTF-8 characters in SMTP envelope
   and mail headers. To deliver Non-ASCII mail address through EAI
   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 systems which are defined by [RFC2821] and [RFC2822]
allow US-ASCII characters in SMTP envelope and mail headers in body
part.   The EAI proposal [EAI-Overview],[EAI-UTF8], [EAI-SMTPext]
allows UTF-8 characters in SMTP envelope and mail headers in body
part.

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

Downgrading in EAI consists from following two parts:
o SMTP session downgrading
o header downgrading

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 downgrading is described in
Section 4, and mail header downgrading is described in Section 5.

2. Terminology

Terminology for this document is defined in [EAI-Overview].

In this document, "algorithmic address" is an US-ASCII address which
is generated by algorithmic method.

3. Downgrade Requirements

3.1. Timing and conditions of downgrading

This section describes timing and conditions of downgrading.
o Timing: SMTP client detects that SMTP server doesn't support
   "IEmail" option at EHLO. [EAI-SMTPext]
o Conditions: SMTP client detects that UTF-8 is included in the SMTP
   envelope or mail headers in the SMTP DATA.

Note: If the i-Email header exists, downgrading will be performed.
If UTF-8 characters exist in mail headers without the i-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 upgrading must be automated.
4.   Downgrading and upgrading 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 upgrading 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:

o   MAIL FROM:
o   RCPT TO:

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

Downgrading is possible only when a mail sender's MUA appends ALT-
ADDR or ATOMIC option to all Non-ASCII envelope addresses to denote
their alternative US-ASCII address.

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

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

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

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

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

MTA replaces Non-ASCII mail address with specified or generated
alternative US-ASCII address. Then appends replaced information with
EAI-Downgraded-From and EAI-Downgraded-To header in mail header
(outgoing SMTP DATA).
    EAI-Downgraded-From: <Non-ASCII,ATOMIC> <US-ASCII>
    EAI-Downgraded-From: <Non-ASCII,US-ASCII> <US-ASCII>
    EAI-Downgraded-To: <Non-ASCII,ATOMIC> <US-ASCII>
    EAI-Downgraded-To: <Non-ASCII,US-ASCII> <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 EAI-Downgraded-From/To headers, MUA/MTA
MUST perform SMTP DATA/Header downgrading. This is described in next
section.

Downgraded local-part is parsed only in MDA. MDA delivers the mail
to final mailbox.

Case study: SPF check

SPF checks domainname of the envelope from and smtp connection IP
address. If ALT-ADDR domainname is Punycode/IDNA form of Non-ASCII
domainname, it will be compatible with current SPF. In this case,
SPF check will be performed correctly. Otherwise, more detailed
consideration is required.

5. SMTP DATA/Header downgrading

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

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

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

Originator address(es): Non-ASCII mail addresses in From, Reply-To,
   Sender and their Resent- headers MUST be target of downgrading.
Destination address(es): Non-ASCII mail addresses 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 Non-ASCII mail
   addresses 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.

Pros:
   * Easy to implement.
Cons:
   * This method may break existing mail infrastructure.

5.2. Downgrading with MIME encapsulation

This downgrading method requires new MIME 'Content-Type:' which
express EAI. 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).
   * Message-ID, Subject, Date headers are copied from original
      header.
   * From header is generated with downgraded Envelope-from.
   * To header is generated with single downgraded Envelope-to.
   * If Subject header contains UTF-8, it is replaced to a certain
      message or encoded by MIME [RFC2047].
   * Message-ID, Date headers are preserved.
   As a result, new body contains one new MIME part (Message/EAI).

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 it is a downgraded message, compare mail body's
    message-id and MIME part's message-id. If message-ids are the
    same, it is a 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 upgrading the downgraded message first.

Pros:
*   MTA does not need to decode each header carefully.
*   Whole headers can be submitted AS IS.
Cons:
*   Non-ASCII from/to can not distinguish from downgraded mail
    headers.
*   EAI incompliant MUA can not treat any downgraded mail.

[[Reference to [EAI-Scenarios] and evaluation of each case should be
described here.]]

5.2.1. Downgrading with MIME encapsulation example

    Downgrading 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: <US-ASCII_FROM>
    To: <US-ASCII_TO>
    Date: DATE

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

    EAI-Downgraded-From: <Non-ASCII,ATOMIC> <US-ASCII_FROM>
    EAI-Downgraded-To: <Non-ASCII,ATOMIC> <US-ASCII_TO>
    Received: ...
    Received: ...
    Message-Id: MESSAGE_ID
    Mime-Version: 1.0
    Subject: UTF-8_SUBJECT
    From: <Non-ASCII,ATOMIC>
    To: <Non-ASCII,ATOMIC>
    Date: DATE

    MAIL_BODY

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

5.3. Header conversion

    Define conversion method to US-ASCII for each header which may
    contain Non-ASCII characters. Each header has its own downgrading
    method.

    To preserve all header information, define generic encapsulation
    header: "Downgraded: HeaderName: HeaderValue". The header value is
    encoded by [RFC2047] with UTF-8 tag.

    Downgrading:
       * For all headers, check if the header contains UTF-8 characters.
       * Encapsulate 'i-Email' header in Downgraded header.

       * If the header contains UTF-8 characters,
          + If the header is an address header which is described in
             Section 5.3.1,
             - Preserve the header in 'Downgraded' header.
             - Downgrade the header defined in Section 5.3.1.
          + The other header case, encode the header by [RFC2047] with
             UTF-8 tag.
    Upgrading:
       * If the mail has 'Downgraded' headers, the mail is a downgraded
          EAI mail message.
       * Decode all 'Downgraded' header.
          + Decode header value field string which is [RFC2047] encoded.
          + If the header is address headers described in Section 5.3.1,
             - Apply address header downgrading to the decoded header.
             - Remove the header line which is same to the downgraded
                line.
          + Remove the 'Downgraded' header.
          + Add decoded header to mail header. "HeaderName:
             HeaderValue".
       * If each mail header has [RFC2047] encoded part and which
          encoding is "UTF-8", it is a downgraded header, so decode it.

    Pros:
       * EAI incompliant MUA displays the downgraded mail body except
          original Non-ASCII mail addresses.
       * EAI incompliant MUA displays and handles the sender specified
          or algorithmic address.
       * EAI compliant MUA displays and handles original headers.
    Cons:
       * Implementation and processing cost is higher than 'Header
          Encapsulation' defined in Section 5.2 because MUA/MTA must
          parse each header and encode it by defined method.
       * Hard to preserve whole information AS IS. The address headers
          are preserved but the other headers which is [RFC2047] encoded
          with UTF-8 tag are not distinguished that it is downgraded or
          it is encoded by sender's MUA. Therefore, to check DKIM
          requires special consideration.

    [[Reference to [EAI-Scenarios] and evaluation of each case should be
    described here.]]

5.3.1. Downgrading address headers

    This section targets From, Sender, Reply-To, To, CC, BCC, Resent-
    From, Resent-To, Resent-CC, Resent-Bcc, Resent-sender headers which
    contains Originator/Destination address(es).

    The header value is composed of single or multiple mailbox/angle-addr

   fields defined in [EAI-UTF8].

   If the header contains UTF-8 characters, downgrading method is
   follows.
   1. Extract every field and downgrade mailbox/angle-addr described
       below.
   2. By mailbox/angle-addr downgrading, if the field became empty, the
       field should be removed.
   3. If all header field is removed, remove the header.
   4. If From header is removed, generate new From header from
       envelope-from address.

   EAI angle-addr defined in [EAI-UTF8] consists of 4 forms.
   Downgrading method is defined for each form.
   1. <Non-ASCII>
           Non-ASCII mail address without ALT-ADDR and ATOMIC parameter
           case, remove this angle-addr.
   2. <Non-ASCII,US-ASCII>
           Non-ASCII mail address with sender-specified US-ASCII address
           case, replace it as <US-ASCII>.
   3. <Non-ASCII,ATOMIC>
           Non-ASCII mail address with ATOMIC parameter case, generate
           the algorithmic address from Non-ASCII mail address and
           replace it as <ALG-ASCII>.
   4. <US-ASCII>
           US-ASCII mail address case, preserve it.

   "mailbox" is defined as "DISPLAY NAME angle-addr" in [EAI-UTF8]. The
   "DISPLAY NAME" field should be encoded by [RFC2047] with UTF-8 tag,
   if necessary. If the angle-addr is removed, remove the field
   including "DISPLAY NAME".

5.3.2. Header conversion example

Original EAI message

```
i-Email: 1.0
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Content-Type: text/plain; charset="UTF-8"
Content-Transfer-Encoding: 8bit
Subject: UTF-8_SUBJECT
From: <NON-ASCII-FROM,ASCII-FROM>
To: <NON-ASCII-TO,ASCII-TO>
CC: <NON-ASCII-CC,ASCII-CC>
Date: DATE

MAIL_BODY
```

SMTP downgrading adds EAI-Downgraded-From, EAI-Downgraded-To headers.

```
EAI-Downgraded-From: <Non-ASCII,DOWNGRADED_FROM> <DOWNGRADED_FROM>
EAI-Downgraded-To: <Non-ASCII,DOWNGRADED_TO> <DOWNGRADED_TO>
```

Result of the header conversion downgrading.

```
EAI-Downgraded-From:
    MIME(<Non-ASCII,DOWNGRADED_FROM>) <DOWNGRADED_FROM>
EAI-Downgraded-To:
    MIME(<Non-ASCII,DOWNGRADED_TO>) <DOWNGRADED_TO>
Downgraded: i-Email: 1.0
Message-Id: MESSAGE_ID
Mime-Version: 1.0
Content-Type: text/plain; charset="UTF-8"
Content-Transfer-Encoding: 8bit
Subject: MIME(UTF-8_SUBJECT)
Downgraded: From: MIME(<NON-ASCII-FROM,ASCII-FROM>)
From: <ASCII-FROM>
Downgraded: To: MIME(<NON-ASCII-TO,ASCII-TO>)
To: <ASCII-TO>
Downgraded: CC: MIME(<NON-ASCII-CC,ASCII-CC>)
CC: <ASCII-CC>
Date: DATE

MAIL_BODY
```

MIME() stands for [RFC2047] encoding.

6. Implementation consideration

6.1. MUA

EAI compliant MUA MUST implement downgrading mechanism for sending.

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

EAI compliant MUA MUST upgrade downgraded mail and MUST show Non-ASCII mail addresses on display.

6.2. MDA Requirements

This section describes downgrading in MDA.
1. MDA MUST NOT upgrade.
2. Perform downgrading for each Storage/Back-end-Process. If and only if MDA knows recipient's MUA is EAI compliant, then no downgrading is performed.
3. If MDA detects that SMTP recipient address is an algorithmic address, then MDA MUST decode it and perform the same processing as if it were Non-ASCII mail address. MDA MAY normalize or canonicalize local-part before processing it.

7. Security considerations

See the extended security considerations discussion in [EAI-Overview]

8. IANA Considerations

To distinguish downgraded Non-ASCII mail addresses in ACE form, it MUST have ACE-Prefix. The ACE-Prefix MUST differ from IDNA ACE-Prefix to avoid possible confusion. IANA will assign Non-ASCII mail address 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

[EAI-Overview]
            Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", draft-ietf-eai-framework-01

                (work in progress).

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

     [EAI-Scenarios]
                Alvestrand, H., "Internationalized Email Addresses:
                Scenarios", draft-ietf-eai-scenarios-00 (work in
                progress), May 2006.

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

     [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions)
                Part Three: Message Header Extensions for Non-ASCII Text",
                RFC 2047, November 1996.

     [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.

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