

**Internationalized Email Headers**  
**draft-yeh-ima-utf8headers-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 August 31, 2006.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

Full internationalization of electronic mail requires not only the capability to transmit non-ASCII content, to encode selected information in specific header fields, and to use international characters in envelope addresses. It also requires being able to express those addresses and information based on them in mail header fields. This document specifies the use of Unicode encoded in UTF-8, rather than ASCII, as the base form for Internet email header fields. This form is permitted in transmission only if authorized by an SMTP extension, as specified in an associated specification.

Table of Contents

- [1. Introduction . . . . .](#) [3](#)
- [1.1. Role of this specification . . . . .](#) [3](#)
- [2. Background and History . . . . .](#) [3](#)
- [3. Terminology . . . . .](#) [4](#)
- [4. Pre-requirement . . . . .](#) [4](#)
- [5. Identification of internationalized email . . . . .](#) [5](#)
- [6. Impact on Message Header Fields . . . . .](#) [6](#)
- [7. Additional issue . . . . .](#) [7](#)
- [7.1. POP3/IMAP . . . . .](#) [7](#)
- [7.2. Mailing list header fields . . . . .](#) [7](#)
- [7.3. URI/IRI . . . . .](#) [7](#)
- [8. Security Considerations . . . . .](#) [7](#)
- [9. IANA considerations . . . . .](#) [8](#)
- [10. Acknowledgements . . . . .](#) [8](#)
- [11. References . . . . .](#) [8](#)
- [11.1. Normative References . . . . .](#) [8](#)
- [11.2. Informative References . . . . .](#) [9](#)
- [Author's Address . . . . .](#) [10](#)
- [Intellectual Property and Copyright Statements . . . . .](#) [11](#)

Yeh

Expires August 31, 2006

[Page 2]

## **1. Introduction**

### **1.1. Role of this specification**

Full internationalization of electronic mail requires several capabilities:

- o The capability to transmit non-ASCII content, provided for as part of the basic MIME specification [[RFC2045](#)], [[RFC2046](#)].
- o The capability to encode selected information in specific header fields, provided for as another part of the MIME specification [[RFC2047](#)].
- o The capability to use international characters in envelope addresses, discussed in [[IMA-overview](#)] and specified in [IMA-SMTP-extension]. And, finally,
- o The capability to express those addresses, and information related to and based on them, in mail header fields, defined in this document.

This document specifies the use of Unicode encoded in UTF-8 [[RFC3629](#)], rather than ASCII, as the base form for Internet email header fields. This form is permitted in transmission, if authorized by the SMTP extension specified in [[IMA-SMTP-extension](#)].

## **2. Background and History**

Mailbox names often represent the names of human users. Many of these users throughout the world have names that are not normally represented with just the ASCII repertoire of characters, and would more the less like to use their real names in their mailbox names. These users are also likely to use non-ASCII text in their common names and subjects of email messages, both in what they send and what they receive. This protocol specifies UTF-8 as the encoding to represent email header messages.

The traditional format of email messages [[RFC2822](#)] only allows ASCII characters in the header fields of messages. This prevents users from having email addresses that contain non-ASCII characters. It further forces non-ASCII text in common names, comments, and in free text (such as in the Subject: field) to be in MIME format [[RFC2047](#)]. This specification describes a change to the email message format that is connected to the SMTP message transport change described in the associated specifications [[IMA-overview](#)] and [IMA-SMTP-extension], and that allows non-ASCII characters throughout email header fields. These changes affect SMTP clients, SMTP servers, and mail user agents (MUAs).



As specified in [[IMA-SMTP-extension](#)], an SMTP protocol extension [[RFC2821](#)] is used to prevent the transmission of messages with UTF-8 header fields to systems that cannot handle such messages.

Use this SMTP extension helps prevent against the introduction of such messages into message stores that might misrepresent or mangle such messages. It should be noted that using an ESMTP extension does not prevent against transferring email messages with UTF-8 header fields to other systems that use the email format for messages and that may not be upgraded, such as the POP and IMAP protocols. Those protocols will need to be changed in order to handle stored messages that have UTF-8 header fields.

The objective for this protocol is to allow UTF-8 in email header fields. Issues about how to handle messages that contain UTF-8 header fields but are proposed to be delivered to systems that have not been upgraded to support this capability are discussed elsewhere, particularly in [[IMA-downgrading](#)].

This protocol is workable even if IMA mailbox names are not presented. For example, the protocol might still be used if just the subject header has non-ASCII characters, but the protocol MUST be used if other header fields (particularly trace header fields such as "Received:") contain non-ASCII characters.

### **3. Terminology**

In this document, header fields are "UTF-8 header" if the bodies of headers contain UTF-8 characters.

Unless otherwise noted, all terms used here are defined in [[RFC2821](#)] or [[RFC2822](#)] or in [[IMA-overview](#)].

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

This document is being discussed on the ima mailing list. See <https://www1.ietf.org/mailman/listinfo/ima> for information about subscribing. The list's archive is at <http://www1.ietf.org/mail-archive/web/ima/index.html>.

### **4. Pre-requirement**

The use of UTF-8 header fields is dependent on the use of an SMTP extension named "IMA".

Yeh

Expires August 31, 2006

[Page 4]

That protocol is defined in [[IMA-SMTP-extension](#)]. If that extension is not supported, UTF-8 header fields MUST NOT be transmitted.

Sending MUAs that follow this protocol MUST create all header fields encoded in UTF-8. No other direct encodings are allowed. MUAs MAY continue to use MIME to specify some text in other encodings; however this is not recommended because it is likely that this will not interoperate well with MUAs that follow this specification.

## 5. Identification of internationalized email

When a SMTP client tries to send a mail to a SMTP server that does not support IMA, the client should know whether the message requires the support for IMA or not. In addition to this, identification of internationalized email is also required when a message is stored and presented. Checking the presence of UTF-8 characters in the header whenever such an identification is required may also achieve the its goal. However, this type of repeated processing wastes time and processing power of involved systems. It is nice to have a mechanism (such as self-label) or some indicator to identify whether the message is new format(i.e. IMA compliant) or old one (i.e. [RFC 2822](#) compliant).

To be able to do so, sending MUA should insert a new header field to identify the presence of i18n information (particularly UTF-8 headers) in the message. The new header specified as "i18n-email", and elements of the header is the version number of i18n email. The i18n header field syntax specified like:

```
i18n-email: 1.0
```

[Note in draft: There should be more useful information can be place in the new header field. ]

While we can't require ordering of headers, it would be good to have it appear as near the top of the headers as possible. It would also be good to be able to guarantee that it will be there when the message is dropped into a mail store. Thus, when a i18n email is delivered.

- o The "i18n-email" header field MUST be inserted by the originating MUA.
- o The "i18n-email" header field MUST be inserted, along with Return-path, by the final delivery MTA if not presented.
- o The "i18n-email" header field, if present, MUST be removed as part of any downgrading process that eliminates the UTF-8 header information.





- o MTAs MAY check for duplicates of the "i18n-email" header field and eliminate all but one of them. However, if a receiving MUA encounters more than one of these headers, it SHOULD simply ignore any excess ones.

This combination guarantees that the header will be present on delivery even if it is deleted in transit.

## 6. Impact on Message Header Fields

This protocol does NOT change the definition of header field names. That is, only the bodies of header fields are allowed to have UTF-8 characters; the rules in [RFC 2822](#) for header names are not changed.

SMTP client can send header fields in UTF-8 format, if the IEmail extension advertised by SMTP server. However, the Message-ID is the unique identifier of a single email. [Note in draft: Extension name depends on the SMTP extension defined in [IMA-SMTP-extension]] In order to maintain the identity, message identifiers of the Message-ID fields MUST be created in all ASCII.

To be specific, when IEE smtp extension is advertised.

- o <unstructured>, <comment> and <phrase> are allowed to use UTF-8.
- o <date-time>, <msg-id> remains the same definition as in [RFC2822](#).

In this specification, internationalized email address will be presented in UTF-8. Thus, all header fields involving <mailbox>es may be different from traditional ones. There might be IMA unaware MTAs in the mail routing path. In that case, MTA may bounce the message with reply code 558, or downgrade the non-ASCII contents of all header bodies before continuing to send the message, as described in [[IMA-downgrading](#)]. However, MTAs never know if there are any data or instructions embedded in the email address. Or there also email addresses do not contain embedded operations. The only one way is to let the mail address owner to tell if the address is ok for downgrade process or not. Hence, the ATOMIC and ALT-ADDRESS options are introduced. The detail of ATOMIC and ALT-ADDRESS options can be found in [[IMA-SMTP-extension](#)]. With these two different cases, there are two possible representation of <mailbox>.

- o ATOMIC:

ATOMIC, it means that the email address can be downgraded safely without damage to the mail delivery. In this case, the <mailbox> syntax remains the same to [RFC2822](#). The only difference is that the <local-part> and <domain> of <addr-spec> allows UTF-8 characters.



- o ALT-ADDRESS:

If user provides an alternative address for the internationalized email address for the mail delivery. The <mailbox> syntax will be

```
mailbox =          new-name-addr / new-addr-spec
new-name-addr =    [display-name] new-angle-addr
new-angle-addr =   [CFWS] "<" new-addr-spec ">" [CFWS]
new-angle-addr =/= obs-angle-addr
new-addr-spec =    [addr-spec] non-ASCII-addr-spec
new-addr-spec =/=  addr-spec
```

In any time, SMTP server can reject with a reply code of 558 whenever ALT-ADDRESS is not provided and downgrade is not feasible.

[Note in draft: The detail ABNF will need to be prepared in this document when proper WG establish.]

## **7. Additional issue**

This section identifies issues that are not covered as part of this set of specifications, but that will need to be considered as part of IEE deployment.

### **7.1. POP3/IMAP**

Receiving MUAs that follow this protocol MUST able to handle email header fields encoded in UTF-8. Which means that the email fetching protocol such as POP3 or IMAP MAY need to be updated.

### **7.2. Mailing list header fields**

All mailing list and mail redistribution related header fields may need further investigation.

### **7.3. URI/IRI**

The mailto schema in URI/IRI may need further investigation.

## **8. Security Considerations**

If a user has a non-ASCII mailbox address and a all-ASCII mailbox address, a digital certificate that identifies that user SHOULD have both addresses in the identity. Having multiple email addresses as identities in a single certificate is already supported in PKIX and OpenPGP.



Because UTF-8 often requires several octets to encode a single character, internationalized local parts may cause mail addresses to become longer. This may possibly make it harder to keep lines in a header under 78 octets. Lines that are longer than 78 octets (which is a SHOULD specification, not a MUST specification, in [RFC 2822](#)) could possibly cause mail user agents to fail in ways that affect security.

## **9. IANA considerations**

The ESMTP extension needed to support this specification is specified in [[IMA-SMTP-extension](#)]. This specification does not require any additional IANA actions in that regard.

## **10. Acknowledgements**

This document was created by incorporating a good deal of material from an old Internet Draft by Paul Hoffman [[Hoffman-utf8-headers](#)]. While many of the concepts and details have changed, the contributions from that draft are greatly appreciated.

Most of the content of this document is provided by John C Klensin. Also some significant comments and suggestions were received from Charles H. Lindsey, Yangwoo KO, Yoshiro YONEYA, and other members of the JET team and were incorporated into the document. The editor is much great thanks to their contribution sincerely.

## **11. References**

### **11.1. Normative References**

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

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

[IMA-SMTP-extension]

Yao, J., Ed. and X. LEE, "SMTP extension for internationalized email address", [draft-yao-ima-smtpext-00.txt](#) (work in progress), January 2006.



[IMA-overview]

Klensin, J. and Y. Ko, "Overview and Framework of Internationalized Email Address Delivery", [draft-klensin-ima-framework-00.txt](#) (work in progress), September 2005.

[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.

[RFC3066] Alvestrand, H., "Tags for the Identification of Languages", [BCP 47](#), [RFC 3066](#), January 2001.

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

## **11.2. Informative References**

[Hoffman-utf8-headers]

Hoffman, P., "SMTP Service Extensions or Transmission of Headers in UTF-8 Encoding", [draft-hoffman-utf8headers-00.txt](#) (work in progress), December 2003.

[IMA-downgrading]

"whatever we call the downgrading document", 2005.

[RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", [RFC 2045](#), November 1996.

[RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", [RFC 2046](#), November 1996.

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





Author's Address

Jeff Yeh (editor)  
TWNIC  
4F-2, No. 9, Sec 2, Roosevelt Rd.  
Taipei, 100  
Taiwan

Phone: +886 2 23411313 ext 506

Email: [jeff@twmic.net.tw](mailto:jeff@twmic.net.tw)

## Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

## Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

## Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

## Acknowledgment

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

