

Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: March 6, 2008

J. Yao, Ed.  
W. Mao, Ed.  
CNNIC  
September 3, 2007

**SMTP extension for internationalized email address  
draft-ietf-eai-smtpext-08.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 March 6, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2007).

Abstract

This document specifies the use of SMTP extension for internationalized email address delivery. Communication with systems that do not implement this specification is specified in another document.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">1.1.</a>	Role of this specification . . . . .	<a href="#">3</a>
<a href="#">1.2.</a>	Proposal Context . . . . .	<a href="#">3</a>
<a href="#">1.3.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Mail Transport-level Protocol . . . . .	<a href="#">4</a>
<a href="#">2.1.</a>	Framework for the Internationalization Extension . . . . .	<a href="#">4</a>
<a href="#">2.2.</a>	The UTF8SMTP Extension . . . . .	<a href="#">5</a>
<a href="#">2.3.</a>	Extended Mailbox Address Syntax . . . . .	<a href="#">6</a>
<a href="#">2.4.</a>	The ALT-ADDRESS parameter . . . . .	<a href="#">8</a>
<a href="#">2.5.</a>	ALT-ADDRESS parameter usage and response codes . . . . .	<a href="#">9</a>
<a href="#">2.6.</a>	Body Parts and SMTP Extensions . . . . .	<a href="#">9</a>
<a href="#">2.7.</a>	Additional ESMTP Changes and Clarifications . . . . .	<a href="#">10</a>
<a href="#">2.7.1.</a>	The Initial SMTP Exchange . . . . .	<a href="#">10</a>
<a href="#">2.7.2.</a>	Mail eXchangers . . . . .	<a href="#">10</a>
<a href="#">2.7.3.</a>	Trace Information . . . . .	<a href="#">10</a>
<a href="#">2.7.4.</a>	UTF-8 Reply . . . . .	<a href="#">11</a>
<a href="#">3.</a>	Issues with Other Parts of the Email System . . . . .	<a href="#">13</a>
<a href="#">3.1.</a>	LMTP . . . . .	<a href="#">13</a>
<a href="#">3.2.</a>	SMTP Service Extension for DSNs . . . . .	<a href="#">13</a>
<a href="#">3.3.</a>	POP and IMAP . . . . .	<a href="#">13</a>
<a href="#">4.</a>	Potential problems . . . . .	<a href="#">14</a>
<a href="#">4.1.</a>	Impact many email related RFC . . . . .	<a href="#">14</a>
<a href="#">5.</a>	Implementation Advice . . . . .	<a href="#">14</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">14</a>
<a href="#">7.</a>	Security considerations . . . . .	<a href="#">15</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">15</a>
<a href="#">9.</a>	Change History . . . . .	<a href="#">15</a>
<a href="#">9.1.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 00 . . . . .	<a href="#">15</a>
<a href="#">9.2.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 01 . . . . .	<a href="#">15</a>
<a href="#">9.3.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 02 . . . . .	<a href="#">16</a>
<a href="#">9.4.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 03 . . . . .	<a href="#">16</a>
<a href="#">9.5.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 04 . . . . .	<a href="#">16</a>
<a href="#">9.6.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 05 . . . . .	<a href="#">16</a>
<a href="#">9.7.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 06 . . . . .	<a href="#">16</a>
<a href="#">9.8.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 07 . . . . .	<a href="#">16</a>
<a href="#">9.9.</a>	<a href="#">draft-ietf-eai-smtpext</a> : Version 08 . . . . .	<a href="#">17</a>
<a href="#">10.</a>	References . . . . .	<a href="#">17</a>
<a href="#">10.1.</a>	Normative References . . . . .	<a href="#">17</a>
<a href="#">10.2.</a>	Informative References . . . . .	<a href="#">18</a>
	Authors' Addresses . . . . .	<a href="#">19</a>
	Intellectual Property and Copyright Statements . . . . .	<a href="#">20</a>



## **1. Introduction**

Internationalized email address includes two parts, the local part and the domain part. The ways email addresses are used by protocols are different from the ways domain names are used. The most critical difference is that emails are delivered through a chain of peering clients and servers while domain names are resolved by name servers by looking up their own tables. In addition to this, email transport protocol ESMTP[RFC1869] provides a negotiation mechanism through which clients can make decisions for further processing; please see more in [[EAI-framework](#)]. Email addresses can exploit the SMTP extension negotiation mechanism while Internationalized Domain Name(IDN) does not have such a facility. This is also more architecturally desirable approach. This document specifies an SMTP extension to permit internationalized email addresses in envelopes, and UTF-8 in headers. The protocol described here is an MTA solution which is feasible, architecturally elegant, and not difficult to deploy.

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

The framework document [[EAI-framework](#)] specifies the requirements for, and components of, full internationalization of electronic mail. To understand and implement this specification, understanding the context presented in [[EAI-framework](#)] is necessary.

This document specifies an element of that work, specifically the definition of an SMTP extension [[RFC1869](#)] for the internationalized email address transport delivery.

### **1.2. Proposal Context**

This specification describes a change to the email transport mechanism that permits non-ASCII characters in both the envelope and header fields of messages while the specification in [[EAI-utf8header](#)] specifies the details of how and where non-ASCII characters are permitted in the header fields of messages. The context for the change is described in [[EAI-framework](#)].

### **1.3. Terminology**

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

All specialized terms used in this specification are defined in the EAI framework [[EAI-framework](#)] or in [[RFC2821](#)] and [[RFC2822](#)]. The terms "ASCII address", "internationalized email address", "non-ASCII



address", "i18mail address", "UTF8SMTP", "message" and "mailing list" are used with the definitions from the [[EAI-framework](#)] document.

This document defines only those ABNF [[RFC4234](#)] syntax rules that are different from those of the base email specifications [[RFC2821](#)][[RFC2822](#)] and, where the earlier rules are upgraded or extended, gives them new names. When the new rule is a small upgrade to the older one, it is typically given a name starting with "u". Rules that are undefined here may be found in the base email documents under the same names.

[[anchor4: RFC EDITOR'S NOTE: The following text should be deleted before publication.]] This document is being discussed on the EAI 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>.

## **2. Mail Transport-level Protocol**

### **2.1. Framework for the Internationalization Extension**

The following service extension is defined:

1. The name of the SMTP service extension is "Email Address Internationalization";
2. The EHLO keyword value associated with this extension is "UTF8SMTP";
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 that keyword. Clients MUST ignore any parameters, that is, clients MUST behave as if the parameters do not appear. If a server includes UTF8SMTP in its EHLO response, it MUST be fully compliant with this version of this specification.
4. One optional parameter, ALT-ADDRESS, is added to the SMTP MAIL and RCPT commands. ALT-ADDRESS specifies an all-ASCII address which can be used as a substitute for the i18mail addresses that we call the primary address; you can learn more in [[EAI-framework](#)] or [[EAI-downgrading](#)].
5. One optional parameter "UTF8REPLY" is added to the VRFY and EXPN commands. The parameter UTF8REPLY has no value. The parameter indicates the SMTP client can accept UTF-8 on replies of the VRFY and EXPN commands.
6. No additional SMTP verbs are defined by this extension.
7. Servers offering this extension MUST provide support for, and announce, the 8BITMIME extension [[RFC1652](#)].



8. The reverse-path and forward-path of SMTP MAIL and RCPT commands are extended to allow UTF-8 characters in the specified mailbox address.
9. The mail datum is extended in compliance with [\[EAI-utf8header\]](#)
10. The maximum length of a MAIL and RCPT command lines is increased by 460 characters by the possible addition of the ALT-ADDRESS keyword and value.

## **[2.2.](#) The UTF8SMTP Extension**

An SMTP Server that announces this extension MUST be prepared to accept a UTF-8 string [\[RFC3629\]](#) in any position in which [RFC 2821](#) specifies that a "mailbox" MAY appear. That string MUST be parsed only as specified in [RFC 2821](#), i.e., by separating the mailbox into source route, local part and domain part, using only the characters colon (U+003A), comma (U+002C), and at-sign (U+0040) as specified there. Once isolated by this parsing process, the local part MUST be treated as opaque unless the SMTP Server is the final delivery MTA. Any domain names that are to be looked up in the DNS MUST first be processed into the form specified in IDNA [\[RFC3490\]](#) by means of the ToASCII() operation unless they are already in that form. Any domain names that are to be compared to local strings SHOULD be checked for validity and then MUST be compared as specified in [section 3.4](#) of IDNA.

The UTF8SMTP extension is valid on the submission port [\[RFC4409\]](#).

An SMTP Client that receives the UTF8SMTP extension keyword in response to the "EHLO" command MAY transmit a mailbox name as an internationalized string in UTF-8 form and MAY send an UTF-8 header [\[EAI-utf8header\]](#). It MAY transmit the domain part of that string in either the form of ACE labels specified in [\[RFC3490\]](#) or UTF-8 form. In the domain part of the mailbox string, if any of the labels are intended to be interpreted as non-ASCII (i.e., are IDNs), then the Message Submission Server ("MSA") [\[RFC4409\]](#) MUST take responsibility for ensuring that the labels are valid (whether they appear in native character or ACE form). The presence of the UTF8SMTP extension does not change the requirement of [RFC 2821](#) that servers relaying mail MUST not attempt to parse, evaluate, or transform the local part in any way.

If the UTF8SMTP SMTP extension is not offered by the Server, the SMTP client MUST NOT transmit an internationalized address and MUST NOT transmit a mail message which contains internationalized mail headers [\[EAI-utf8header\]](#) at any level within its MIME structure. Instead, if an SMTP client (SMTP sender) attempts to transfer a UTF8SMTP message and encounters a server that does not support the extension, it MUST make one of the following four choices:





1. If and only if the SMTP client (sender) is a Message Submission Server[RFC4409], it MAY, consistent with the general provisions for changes by such servers, rewrite the envelope, headers, or message material to make them entirely ASCII and consistent with the provisions of [RFC 2821](#) [RFC2821] and [RFC 2822](#) [RFC2822].
2. Reject the message during the SMTP transaction or generate a notification of non-deliverability, as specified in [RFC 2821](#) [RFC2821] and [RFC 3464](#) [RFC3464]. If the message content can be returned without alteration, content should be returned as specified in 2821 but, if a server is encountered along the return path that cannot accept UTF8SMTP traffic, the content should simply be abridged or dropped.
3. Find an alternate route to the destination that permits UTF8SMTP. That route may be discovered by trying alternate MX hosts (using preference rules as specified in [RFC 2821](#)) or using other means available to the SMTP-sender.
4. If and only if ASCII addresses are available for all addresses that appear in the return path and the specific forward paths being attempted, downgrade the message to an all-ASCII form as specified in [[EAI-downgrading](#)]. An ASCII address is considered to be "available" for a particular address if the original address in the envelope is in ASCII or if an ALT-Address parameter is specified for a UTF8SMTP address.

### **[2.3.](#) Extended Mailbox Address Syntax**

[RFC 2821, section 4.1.2](#), defines the syntax of a mailbox entirely in terms of ASCII characters, using the production for "Mailbox" and those on which it depends.

The key changes made by this specification are, informally, to

- o Change the definition of "sub-domain" to permit either the definition above or a UTF-8 string representing a DNS label that is conformant with IDNA [[RFC3490](#)].
- o Change the definition of "Atom" to permit either the definition above or a UTF-8 string. That string MUST NOT contain any of the ASCII characters (either graphics or controls) that are not permitted in "atext"; it is otherwise unrestricted.

According to the description above, define the syntax of an internationalized email mailbox with ABNF [[RFC4234](#)] as



uMailbox = uLocal-part "@" uDomain  
; Replace Mailbox in [RFC 2821, section 4.1.2](#)

uLocal-part = uDot-string / uQuoted-string  
; MAY be case-sensitive  
; Replace Local-part in [RFC 2821, section 4.1.2](#)

uDot-string = uAtom \*("." uAtom)  
; Replace Dot-string in [RFC 2821, section 4.1.2](#)

uAtom = 1\*ucharakter  
; Replace Atom in [RFC 2821, section 4.1.2](#)

ucharakter = atext / UTF8-xtra-char  
; Replace character in [RFC 2821, section 4.1.2](#)  
; atext is defined in [RFC 2822](#)

uQuoted-string = DQUOTE \*uqcontent DQUOTE  
; Replace Quoted-string in [RFC 2821, section 4.1.2](#)  
; DQUOTE is Double Quote defined in [RFC 4234](#)

uqcontent = qcontent / UTF8-xtra-char  
; qcontent is defined in [RFC 2822, section 3.2.5](#)

uDomain = (sub-udomain 1\*("." sub-udomain)) / address-literal  
; Replace Domain in [RFC 2821, section 4.1.2](#)  
; address-literal is defined in [RFC2821 section 4.1.2](#)

sub-udomain = uLet-dig [uLdh-str]  
; Replace sub-domain in [RFC 2821, section 4.1.2](#)

uLet-dig = Let-dig / UTF8-xtra-char  
; Let-dig is defined in [RFC 2821, section 4.1.3](#)

uLdh-str = \*( ALPHA / DIGIT / "-" / UTF8-xtra-char ) uLet-dig  
; Replace Ldh-str in [RFC 2821, section 4.1.3](#)

UTF8-xtra-char = UTF8-2 / UTF8-3 / UTF8-4  
; UTF8-2, UTF8-3 and UTF8-4 are defined in [RFC 3629](#)

The value of "udomain" SHOULD be verified with IDNA [[RFC3490](#)]; If failed, the email address with that udomain can not be regarded as the valid email address.



#### **2.4. The ALT-ADDRESS parameter**

If the UTF8SMTP extension is offered, the syntax of the SMTP MAIL and RCPT commands is extended to support the optional esmtp-keyword "ALT-ADDRESS", which specifies an alternate all-ASCII address which may be used when downgrading. If the ALT-ADDRESS esmtp-keyword is used, it MUST have an associated esmtp-value (ALT-ADDRESS-esmtp-value which is defined below).

Based on the definition of mail-parameters in [[RFC2821](#)], the ALT-ADDRESS parameter usage in the commands of "MAIL" and "RCPT" is defined below.

```
"MAIL FROM:" ("<" / uReverse-path) [ SP Mail-parameters ] CRLF
; Update mail command in RFC 2821, section 4.1.1.2.
; A new parameter defined by the ABNF non-terminal
; <ALT-ADDRESS-parameter> is added. It complies
; with the syntax specified by <esmtp-param>.
```

```
"RCPT TO:" ("<Postmaster@" domain ">" / "<Postmaster>" /
uForward-Path) [ SP Rcpt-parameters ] CRLF
; Update rcpt command in RFC 2821, section 4.1.1.3.
; A new parameter defined by the ABNF non-terminal
; <ALT-ADDRESS-parameter> is added. It complies
; with the syntax specified by <esmtp-param>.
```

```
uReverse-path = uPath
; Replace Reverse-path in RFC 2821, section 4.1.2
```

```
uForward-path = uPath
; Replace Forward-path in RFC 2821, section 4.1.2
```

```
uPath = "<" [ A-d-l ":" ] uMailbox ">"
; Replace Path in RFC 2821, section 4.1.2
; A-d-l is defined in RFC 2821, section 4.1.2
; uMailbox is defined in section 2.3 of this document
```

```
ALT-ADDRESS-parameter="ALT-ADDRESS=" ALT-ADDRESS-esmtp-value
```

```
ALT-ADDRESS-esmtp-value=xtext
; Mailbox encoded as xtext.
; xtext is defined in RFC 3461 \[RFC3461\], section 4.2
```

The ALT-ADDRESS-parameter MUST NOT appear more than once in any MAIL or RCPT command. ALT-ADDRESS-esmtp-value MUST be an all-ASCII email



address before xtext encoding.

## 2.5. ALT-ADDRESS parameter usage and response codes

[EAI-utf8header] specifies "UTF8SMTP message" which requires UTF8SMTP support. Such a message MUST NOT be sent to an SMTP server which does not support UTF8SMTP. Such a message MAY be rejected due to lack of the ALT-ADDRESS as discussed in [section 2.2](#) of this document.

When messages are rejected because they require UTF8SMTP, the response code "550" is used, defined in [\[RFC2821\]](#), meaning "mailbox unavailable". If enhanced mail system status codes [\[RFC3463\]](#) are used, the response code should be "5.6.x" [\[SMTP-codes\]](#), meaning that "The alt-address is required but not specified".

If the response code is issued after the final "." of the DATA command, the response code "554" is used, defined in [\[RFC2821\]](#), meaning "Transaction failed". If enhanced mail system status codes [\[RFC3463\]](#) are used, the response code should be "5.6.z" [\[SMTP-codes\]](#), meaning that "UTF8SMTP downgrade failed".

[[anchor8: REMOVE THIS: IANA please assign the proper error codes for "5.6.x" and "5.6.z".]]

## 2.6. Body Parts and SMTP Extensions

Since there is no ESMTP parameter which tells whether the message is UTF8SMTP message, SMTP server needs to parse all message header fields and MIME header fields in the message body to discover which messages are UTF8SMTP. While this specification requires that servers support the 8BITMIME extension [\[RFC1652\]](#) 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 addresses obviously does not require non-ASCII body parts in the MIME message. The UTF8SMTP extension MAY be used with the BODY=8BITMIME parameter if that is appropriate given the body content or, if the server advertises it and it is appropriate, with the BODY=BINARYMIME parameter specified in [\[RFC3030\]](#).

Assuming that the server advertises UTF8SMTP and 8BITMIME, and at least one non-ASCII address, with or without ALT-ADDRESS, the precise interpretation of "No 'Body' parameter", "BODY= 8BITMIME", and "BODY= BINARYMIME" in the MAIL command is:

1. For No "Body" parameter, headers are in UTF-8, body parts are in ASCII.
2. For BODY=8BITMIME parameter, headers are in UTF-8, some or all body parts contain 8-bit line-oriented data.





3. For BODY=BINARYMIME parameter, headers are in UTF-8, some or all body parts contain binary data without restriction as to line lengths or delimiters.

## **2.7. Additional ESMTP Changes and Clarifications**

The mail transport process involves addresses ("mailboxes") and domain names in contexts in addition to the MAIL and RCPT commands and extended alternatives to them. In general, the rule is that, when [RFC 2821](#) specifies a mailbox, this document expects UTF-8 to be used for the entire string; when [RFC 2821](#) specifies a domain name, the name SHOULD be in the form of ACE labels if its raw form is non-ASCII.

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

Support and use of this extension requires support for 8BITMIME. It means that 8BITMIME MUST be advertised by the UTF8SMTP capable SMTP server.

### **2.7.1. The Initial SMTP Exchange**

When an SMTP or ESMTP connection is opened, the server normally sends a "greeting" response consisting of the '220' reply code and some information. The client then sends the EHLO command. Since the client cannot know whether the server supports UTF8SMTP until after it receives the response from EHLO, any domain names that appear in this dialogue, or in responses to EHLO, MUST be in the hostname form, i.e., internationalized ones MUST be in the form of ACE labels.

### **2.7.2. Mail eXchangers**

Commonly, organizations authorize multiple servers to accept mail addressed to them. For example, the organization may itself operate more than one server, and may also or instead have an agreement with other organizations to accept mail as a backup. Authorized servers are generally listed in MX records [[RFC2821](#)]. When more than one server accepts mail for the domain-part of a Mailbox, it is strongly advised that either all or none of them support the UTF8SMTP extension. Otherwise, surprising downgrades can happen during temporary failures, which is not a good thing.

### **2.7.3. Trace Information**

When an SMTP server receives a message for delivery or further processing, it MUST insert trace ("time stamp" or "Received") information at the beginning of the message content. Time stamp appears in the form of "Received: lines". The most important use of



Received: lines is for debugging mail faults. When the delivery SMTP server makes the "final delivery" of a message, it inserts a return-path line at the beginning of the mail data. The primary purpose of the Return-path is to designate the address to which messages indicating non-delivery or other mail system failures are to be sent. For the trace information, we update the time stamp line and the return path line [\[RFC2821\]](#) formally defined as follows:

```
uReturn-path-line = "Return-Path:" FWS uReverse-path <CRLF>
    ; Replaces Return-path-line in the section 4.4 of \[RFC2821\]
    ; uReverse-path is defined in Section 2.3
```

```
uTime-stamp-line = "Received:" FWS uStamp <CRLF>
    ; Replaces Time-stamp-line in the section 4.4 of \[RFC2821\]
```

```
uStamp = From-domain By-domain uOpt-info ";" FWS date-time
    ; Replaces Stamp in the section 4.4 of \[RFC2821\]
```

```
uOpt-info = [Via] [With] [ID] [uFor]
    ; Replaces Opt-info in the section 4.4 of \[RFC2821\]
    ; [With]'s protocol value will allow UTF8SMTP value
```

```
uFor = "FOR" 1*( FWS (uPath / uMailbox) ) CFWS
    ; Replaces For in the section 4.4 of \[RFC2821\]
    ; uPath is defined in section 2.4 of this document
    ; uMailbox is defined in section 2.3 of this document
```

[[anchor12: Note: Whether the FOR parameter is permitted to accept more than one address is now under discussion as part of the rfc2821bis effort. The multiple-address construction was introduced with [RFC 2821](#); it is not clear that it has been widely implemented or that it is wise. Whatever decision is reached about RFC2821bis will be reflected in the syntax of a future version of this document.]]

Except in the 'uFor' and 'uReverse-path' line where non-ASCII domain name may be used, internationalized domain names in Received fields MUST be transmitted in the form of ACE labels. The protocol value of the WITH clause is UTF8SMTP when this extension is used. More information is in the "IANA Considerations" section of this document.

#### [2.7.4. UTF-8 Reply](#)

If the client issues the RCPT command which contains non-ASCII characters, the SMTP server is permitted to use UTF-8 characters in the email address within 251 and 551 response codes.

If an SMTP client follows this specification and sends any RCPT



commands containing non-ASCII addresses, it MUST be able to accept and process 251 or 551 replies containing UTF-8 email addresses. If a given RCPT command does not include non-ASCII envelope addresses, the 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 addresses.

If the VRFY and EXPN commands have the optional parameter "UTF8REPLY", it indicates the client can accept UTF-8 on replies of the VRFY and EXPN commands. Specially this allows to use UTF-8 on mailboxes and full names which occur on replies. The SMTP client following this specification MUST accept UTF-8 on replies of the VRFY and EXPN commands. However the SMTP server MUST not use UTF-8 on replies, if the SMTP client does not ask UTF-8 replies. Some replies include the mailbox, but usually most of replies do not require that the mailbox is included in it and therefore UTF-8 is not needed. The UTF8REPLY parameter on the VRFY and EXPN commands tells the SMTP server that the client is prepared for UTF-8 on SMTP replies.

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

```
"VRFY" SP (uLocal-part / uMailbox) [SP "UTF8REPLY"] CRLF;  
      ;uLocal-part is defined in section 2.3 of this document  
      ;uMailbox is defined in section 2.3 of this document
```

```
"EXPN" SP (uLocal-part / uMailbox) [SP "UTF8REPLY"] CRLF;  
      ;uLocal-part is defined in section 2.3 of this document  
      ;uMailbox is defined in section 2.3 of this document
```

This parameter "UTF8REPLY" does not have value. If SMTP reply requires UTF-8, but SMTP client does not use "UTF8REPLY" parameter in the VERIFY (VRFY) and EXPAND (EXPN) commands, the response code 252 is used, defined in [[RFC2821](#)], meaning "Cannot VRFY user, but will accept the message and attempt the delivery". Also response code 550 may be used, meaning "Requested action not taken: mailbox unavailable". If enhanced mail system status code [[RFC3463](#)] is used, response codes given on below is used. "UTF8REPLY" on the VERIFY (VRFY) or EXPAND (EXPN) commands enables UTF-8 for that command only.

If a normal (i.e., 250) response 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 <uMailbox>  
      ; uMailbox is defined in section 2.3 of this document  
      ; User Name allows the non-ASCII character.
```

```
uMailbox
```



; uMailbox is defined in [section 2.3](#) of this document

If the SMTP reply requires UTF-8, but UTF-8 is not allowed on reply, and enhanced mail system status code [\[RFC3463\]](#) is used, the response code should be "5.6.y" or "2.6.y" [\[SMTP-codes\]](#), meaning that "The UTF-8 reply required, but not allowed.". [[anchor13: REMOVE THIS: IANA please assign the proper error codes for "5.6.y" and "2.6.y".]]

If the SMTP Client lack of the UTF8SMTP support receives the UTF-8 message on reply, it may crash. UTF-8 messages on reply are only allowed in the commands under the situations described above. Under any other circumstances, UTF-8 messages on the reply MUST NOT be used.

Although UTF-8 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 for any other purposes. SMTP servers MUST NOT include non-ASCII characters except in the limited cases specifically permitted in this section.

### **[3.](#) Issues with Other Parts of the Email System**

#### **[3.1.](#) LMTP**

LMTP [\[RFC2033\]](#) may be used as the final delivery agent. In such cases, LMTP may be arranged to deliver the mail to the mail store. The mail store may not have UTF8SMTP capability. LMTP need to be updated to deal with these situations.

#### **[3.2.](#) SMTP Service Extension for DSNs**

The existing draft standard Delivery status notifications (DSNs)[\[RFC3461\]](#) is presently limited to US-ASCII text in the machine readable portions of the protocol. "International Delivery and Disposition Notifications" [\[EAI-dsn\]](#) adds a new address type for international email addresses so an original recipient address with non-US-ASCII characters can be correctly preserved even after downgrading. If an SMTP server advertises both the UTF8SMTP and the DSN extension, that server MUST implement EAI-dsn [\[EAI-dsn\]](#) including support for the ORCPT parameter.

#### **[3.3.](#) POP and IMAP**

The [\[EAI-framework\]](#) has introduced two documents [\[EAI-pop\]](#) and [\[EAI-imap\]](#) to how to use internationalized user names based on UTF-8 characters for the retrieval of messages from a mail server.





## **4. Potential problems**

### **4.1. Impact many email related RFC**

Internationalized email has implications for all processes and protocols which examine, handle, generate, or otherwise deal with mail. In particular, address parsing or validity checks, message parsing or handling, etc.

## **5. Implementation Advice**

In the absence of this extension, SMTP clients and servers are constrained to using only those addresses permitted by [RFC 2821](#). The local parts of those addresses MAY be made up of any ASCII characters, although some of them MUST be quoted as specified there. It is notable in an internationalization context that there is a long history on some systems of using overstruck ASCII characters (a character, a backspace, and another character) within a quoted string to approximate non-ASCII characters. This form of internationalization SHOULD be phased out as this extension becomes widely deployed but backward-compatibility considerations require that it continue to be supported.

## **6. IANA Considerations**

IANA is requested to add "UTF8SMTP" to the SMTP extensions registry with the entry pointing to this specification for its definition.

IANA is requested to assign the proper error codes "5.6.x", "5.6.z", "5.6.y" and "2.6.y" for this specification based on [\[SMTP-codes\]](#).

The "Mail Transmission Types" registry is requested to be updated to include the following new entries:

WITH protocol types	Description	Reference
-----	-----	-----
UTF8SMTP	UTF8SMTP with Service Extensions	[RFCxxxx]
UTF8SMTPA	UTF8SMTP with SMTP AUTH	[ <a href="#">RFC2554bis</a> ]
		[RFCxxxx]
UTF8SMTPS	UTF8SMTP with STARTTLS	[ <a href="#">RFC3207</a> ]
		[RFCxxxx]
UTF8SMTPSA	UTF8SMTP with both STARTTLS and SMTP AUTH	[ <a href="#">RFC3207</a> ]
		[ <a href="#">RFC2554bis</a> ]
		[RFCxxxx]



[[anchor22: REMOVE THIS: where RFCxxxx represents the future RFC NO. of this document. When this document is published as RFC and assigned with a RFC No., "xxxx" should be replaced with 4-digits No.. "RFC2554bis" should be replaced with the new RFC No. when the "RFC2554bis" document is assigned with the new RFC No.]]

## **7. Security considerations**

See the extended security considerations discussion in [\[EAI-framework\]](#)

## **8. Acknowledgements**

Much of the text in the initial version of this document was derived or copied from [\[Klensin-emailaddr\]](#) with the permission of the author. 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 document. Special thanks to those contributors for this version of the document, those include (but not limited to) 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. Of course, none of the individuals are necessarily responsible for the combination of ideas represented here.

## **9. Change History**

[[anchor25: REMOVE THIS: This section is used for tracking the update of this document. It may be useful to retain parts of it to facilitate establishing dates and documents for the history of this work.]]

### **9.1. [draft-ietf-eai-smtpext](#): Version 00**

This version supercedes [draft-yao-ima-smtpext-03.txt](#). It refines the ABNF definition of the internationalized email address. It represents as the EAI working group document.

### **9.2. [draft-ietf-eai-smtpext](#): Version 01**

- o Upgraded to reflect discussions during IETF 66.
- o Remove the atomic parameter.



- o Add the new section of "the Suggestion of the value of the ALT-ADDRESS parameter".

### **9.3. [draft-ietf-eai-smtpext](#): Version 02**

- o Upgraded to reflect the recent discussion of the ima@ietf.org mailing list.
- o Add the section of "Body Parts and SMTP Extensions".
- o Add the new section of "Change History".
- o Add the subsection about SMTP extensions for DSN.

### **9.4. [draft-ietf-eai-smtpext](#): Version 03**

- o Update the syntax related to mailbox.
- o Update the trace field section.
- o Add the new section about message retry.
- o Update the subsection about SMTP extensions for DSN.

### **9.5. [draft-ietf-eai-smtpext](#): Version 04**

- o Refine some syntax.
- o Delete "Message Header Label" section.
- o Change "bounce" to "reject".

### **9.6. [draft-ietf-eai-smtpext](#): Version 05**

- o Refine the abstract.
- o Delete "The Suggestion of the Value of the ALT-ADDRESS parameter" section.
- o Move original [section 2.7.4](#) and 2.7.5 to [section 3](#) with the name "Issues with other parts of the email system".
- o Add the new section "LMTP".
- o Refine some text according to suggestions from the EAI mailing list discussion
- o Remove the section "Mailing List Question"

### **9.7. [draft-ietf-eai-smtpext](#): Version 06**

- o Delete the section about message retry.
- o Add the new subsection about Mail eXchangers
- o Add the new section about "UTF-8 Reply"
- o Refine some response code for the section "Using the ALT-ADDRESS parameter"

### **9.8. [draft-ietf-eai-smtpext](#): Version 07**



- o Rename the [section 2.5](#)
- o Refine sthe [section 2.7](#)

### **[9.9. draft-ietf-eai-smtpext: Version 08](#)**

- o Refine some texts and update some references

## **[10. References](#)**

### **[10.1. Normative References](#)**

- [ASCII] Cerf, V., "ASCII format for network interchange", [RFC 20](#), October 1969.
- [EAI-framework] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", [RFC 4952](#), July 2007.
- [EAI-utf8header] Abel, Y., "Transmission of Email Headers in UTF-8 Encoding", [draft-ietf-eai-utf8headers-07.txt](#) (work in progress), September 2007.
- [RFC1652] Klensin, J., Freed, N., Rose, M., Stefferud, E., and D. Crocker, "SMTP Service Extension for 8bit-MIMEtransport", [RFC 1652](#), July 1994.
- [RFC1869] Klensin, J., Freed, N., Rose, M., Stefferud, E., and D. Crocker, "SMTP Service Extensions", STD 10, [RFC 1869](#), November 1995.
- [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.
- [RFC3461] Moore, K., "Simple Mail Transfer Protocol (SMTP) Service Extension for Delivery Status Notifications (DSNs)", [RFC 3461](#), January 2003.
- [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.
- [RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", [RFC 3490](#), March 2003.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", [RFC 3629](#), November 2003.
- [RFC4234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 4234](#), October 2005.

## **10.2. Informative References**

- [EAI-downgrading]  
YONEYA, Y., Ed. and K. Fujiwara, Ed., "Downgrading mechanism for Internationalized eMail Address", [draft-ietf-eai-downgrade-04](#) (work in progress), 7 2007.
- [EAI-dsn] Newman, C. and A. Melnikov, "SMTP extensions for DSNS", [draft-ietf-eai-dsn-03.txt](#) (work in progress), 9 2007.
- [EAI-imap]  
Resnick, P. and C. Newman, "Considerations for IMAP in Conjunction with Email Address Internationalization", [draft-ietf-eai-imap-utf8-01](#) (work in progress), March 2007.
- [EAI-mailing list]  
Gellens, R. and E. Chung, "Mailing Lists and Internationalized Email Addresses", [draft-ietf-eai-mailinglist-01.txt](#) (work in progress), January 2007.
- [EAI-pop] Newman, C., "POP3 Support for UTF-8", [draft-ietf-eai-pop-01.txt](#) (work in progress), January 2007.
- [Klensin-emailaddr]  
Klensin, J., "Internationalization of Email Addresses", [draft-klensin-emailaddr-i18n-03](#) (work in progress), July 2005.
- [RFC2033] Myers, J., "Local Mail Transfer Protocol", [RFC 2033](#), October 1996.



## [RFC2554bis]

Siemborski, R. and A. Melnikov, "SMTP Service Extension for Authentication", [draft-siemborski-rfc2554bis-09](#) (work in progress), April 2007.

[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.

[RFC4409] Gellens, R. and J. Klensin, "Message Submission for Mail", [RFC 4409](#), April 2006.

## [SMTP-codes]

Klensin, J., "An IANA Registry for Extended SMTP Status Codes", [draft-klensin-smtp-code-registry-00](#) (work in progress), April 2007.

## Authors' Addresses

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

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

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

Phone: +86 10 58813055  
Email: maowei\_ietf@cnnic.cn



## Full Copyright Statement

Copyright (C) The IETF Trust (2007).

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.

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, THE IETF TRUST 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.

## Intellectual Property

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).

## Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

