Network Working Group
Request for Comments: 5336
Updates: 2821, 2822, 4952
Category: Experimental

J. Yao, Ed.
W. Mao, Ed.
CNNIC

September 2008

SMTP Extension for Internationalized Email Addresses

Status of This Memo

This memo defines an Experimental Protocol for the Internet community. It does not specify an Internet standard of any kind. Discussion and suggestions for improvement are requested. Distribution of this memo is unlimited.

Abstract

This document specifies an SMTP extension for transport and delivery of email messages with internationalized email addresses or header information. Communication with systems that do not implement this specification is specified in another document. This document updates some syntaxes and rules defined in RFC 2821 and RFC 2822, and has some material updating RFC 4952.

Table of Contents

1. Introduction 3
1.1. Role of This Specification 3
1.2. Terminology 3
2. Overview of Operation 4
3. Mail Transport-Level Protocol 4
3.1. Framework for the Internationalization Extension 4
3.2. The UTF8SMTP Extension 5
3.3. Extended Mailbox Address Syntax 7
3.4. The ALT-ADDRESS Parameter 9
3.5. ALT-ADDRESS Parameter Usage and Response Codes 10
3.6. Body Parts and SMTP Extensions 11
3.7. Additional ESMTP Changes and Clarifications 11
3.7.1. The Initial SMTP Exchange 12
3.7.2. Mail eXchangers 12
3.7.3. Trace Information 12
3.7.4. UTF-8 Strings in Replies 13
4. IANA Considerations 15
5. Security Considerations 17
6. Acknowledgements 17
7. References 18
7.1. Normative References 18
7.2. Informative References 19
Appendix A. Material Updating RFC 4952 20
A.1. Conventional Message and Internationalized Message 20
A.2. LMTP 20
A.3. SMTP Service Extension for DSNs 20
A.4. Implementation Advice 20
A.5. Applicability of SMTP Extension to Additional Uses 21

1. Introduction

An 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 clients and servers, while domain names are resolved by name servers looking up those names in their own tables. In addition to this, the Simple Mail Transfer Protocol [RFC2821] provides a negotiation mechanism about service extension with which clients can discover server capabilities and make decisions for further processing. An extended overview of the extension model for internationalized addresses and headers appears in [RFC4952], referred to as "the framework document" or just as "Framework" elsewhere in this specification. This document specifies an SMTP extension to permit internationalized email addresses in envelopes, and UNICODE characters (encoded in UTF-8) [RFC3629] in headers.

1.1. Role of This Specification

The framework document specifies the requirements for, and describes components of, full internationalization of electronic mail. A thorough understanding of the information in that document and in the base Internet email specifications [RFC2821] [RFC2822] is necessary to understand and implement this specification.

This document specifies an element of the email internationalization work, specifically the definition of an SMTP extension [RFC2821] for internationalized email address transport delivery.

1.2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

The terms "conventional message" and "internationalized message" are defined in an appendix to this specification. The terms "UTF-8 string" or "UTF-8 character" are used informally to refer to Unicode characters encoded in UTF-8 [RFC3629]. All other specialized terms used in this specification are defined in the framework document or in the base Internet email specifications [RFC2821] [RFC2822]. In particular, the terms "ASCII address", "internationalized email address", "non-ASCII address", "i18mail address", "UTF8SMTP", "message", and "mailing list" are used in this document according to the definitions in the framework document.

This specification defines only those Augmented BNF (ABNF) [RFC5234] 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 modification 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 specifications under the same names.

2. Overview of Operation

This specification describes an optional extension to the email transport mechanism that permits non-ASCII [ASCII] characters in both the envelope and header fields of messages, which are encoded with UTF-8 [RFC3629] characters. The extension is identified with the token "UTF8SMTP". In order to provide information that may be needed in downgrading, an optional alternate ASCII address may be needed if an SMTP client attempts to transfer an internationalized message and encounters a server that does not support this extension.

The EAI UTF-8 header specification [RFC5335] provides the details of how and where non-ASCII characters are permitted in the header fields of messages. The context for this specification is described in the framework document.

3. Mail Transport-Level Protocol

3.1. 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 MAIL and RCPT commands of SMTP. ALT-ADDRESS specifies an all-ASCII address which can be used as a substitute for the corresponding primary (i18mail) address when downgrading. More discussion of the use of this parameter appears in [RFC4952] and [Downgrade].

5. One optional parameter "UTF8REPLY" is added to the VRFY and EXPN commands. The parameter UTF8REPLY has no value. The parameter indicates that the SMTP client can accept Unicode characters in UTF-8 encoding in replies from 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 the SMTP MAIL and RCPT commands are extended to allow Unicode characters encoded in UTF-8 in mailbox names (addresses).

9. The mail message body is extended as specified in [RFC5335].

10. The maximum length of MAIL and RCPT command lines is increased by 460 characters by the possible addition of the ALT-ADDRESS keyword and value.

11. The UTF8SMTP extension is valid on the submission port [RFC4409].

3.2. 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 can 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 Mail Transfer Agent (MTA). Any domain names that are to be looked up in the DNS MUST first be processed into the form specified in "Internationalizing Domain Names in Applications (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.

An SMTP client that receives the UTF8SMTP extension keyword in response to the EHLO command MAY transmit mailbox names within SMTP commands as internationalized strings in UTF-8 form. It MAY send a UTF-8 header [RFC5335] (which may also include mailbox names in UTF-8). It MAY transmit the domain parts of mailbox names within SMTP commands or the message header as either ACE (ASCII Compatible Encoding) labels (as specified in IDNA [RFC3490]) or UTF-8 strings. All labels in domain parts of mailbox names which are IDNs (either UTF-8 or ACE strings) MUST be valid. If the original client submits a message to a Message Submission Server ("MSA") [RFC4409], it is the responsibility of the MSA that all domain labels are valid; otherwise, it is the original client's responsibility. 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 containing internationalized mail headers as described in [RFC5335] at any level within its MIME structure. (For this paragraph, the internationalized domain name in the form of ACE labels as specified in IDNA [RFC3490] is not considered as "internationalized".) Instead, if an SMTP client (SMTP sender) attempts to transfer an internationalized 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 ("MSA") [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. It may either reject the message during the SMTP transaction or accept the message and then generate and transmit a notification of non-deliverability. Such notification MUST be done as specified in RFC 2821 [RFC2821], RFC 3464 [RFC3464], and the EAI delivery status notification (DSN) specification [RFC5337].

3. It may find an alternate route to the destination that permits UTF8SMTP. That route may be discovered by trying alternate Mail eXchanger (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, it may downgrade the message to an all-ASCII

form as specified in [Downgrade]. 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.

The difference between choice 1 and choice 4 is that choice 1 is constrained by Message Submission [RFC4409], while choice 4 is constrained by [Downgrade].

3.3. 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 a mailbox and those productions on which it depends.

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

- 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].

- 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, the syntax of an internationalized email mailbox name (address) is defined in ABNF [RFC5234] as follows.

```
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*ucharacter
      ; Replace Atom in RFC 2821, Section 4.1.2

ucharacter = atext / UTF8-non-ascii

atext = <See Section 3.2.4 of RFC 2822>

uQuoted-string = DQUOTE *uqcontent DQUOTE
  ; Replace Quoted-string in RFC 2821, Section 4.1.2

DQUOTE = <See appendix B.1 of RFC 5234>

uqcontent = qcontent / UTF8-non-ascii

qcontent = <See Section 3.2.5 of RFC 2822>

uDomain = (sub-udomain 1*("." sub-udomain)) / address-literal
  ; Replace Domain in RFC 2821, Section 4.1.2

address-literal = <See Section 4.1.2 of RFC 2822>

sub-udomain = uLet-dig [uLdh-str]
  ; Replace sub-domain in RFC 2821, Section 4.1.2

uLet-dig = Let-dig / UTF8-non-ascii

Let-dig = <See Section 4.1.3 of RFC 2821>

uLdh-str = *( ALPHA / DIGIT / "-" / UTF8-non-ascii) uLet-dig
  ; Replace Ldh-str in RFC 2821, Section 4.1.3

UTF8-non-ascii = UTF8-2 / UTF8-3 / UTF8-4

UTF8-2 =  <See Section 4 of RFC 3629>

UTF8-3 =  <See Section 4 of RFC 3629>

UTF8-4 =  <See Section 4 of RFC 3629>
```

The value of "uDomain" SHOULD be verified by applying the tests specified as part of IDNA [RFC3490]. If that verification fails, the email address with that uDomain MUST NOT be regarded as a valid email address.

3.4. 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". That keyword specifies an alternate all-ASCII address that 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).

While it may be tempting to consider ALT-ADDRESS as a general-purpose second-chance address, such behavior is not defined here. Instead, in this specification ALT-ADDRESS only has meaning when the associated primary address is non-ASCII and the message is downgraded. This restriction allows for future extension of the specification even though no such extensions are currently anticipated.

Based on the definition of mail-parameters in [RFC2821], the ALT-ADDRESS parameter usage in the commands of MAIL and RCPT is defined as follows. The following definitions are given in the same format as used in RFC 2821.

```
"MAIL FROM:" ("<>" / uReverse-path) [ SP Mail-parameters ] CRLF
   ; Update the 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 for <esmtp-param> in RFC 2821.

"RCPT TO:" ("<Postmaster@" uDomain ">" / "<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 for <esmtp-param>.
       ; uDomain is defined in Section 3.3 of this document.

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.
   ; uMailbox is defined in Section 3.3 of this document.

A-d-l = <See Section 4.1.2 of RFC 2821>

ALT-ADDRESS-parameter = "ALT-ADDRESS=" ALT-ADDRESS-value

ALT-ADDRESS-value = xtext
       ; The value is a mailbox name encoded as xtext.

xtext = <See Section 4.2 of RFC 3461>
```

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.

3.5. ALT-ADDRESS Parameter Usage and Response Codes

An "internationalized message" as defined in the appendix of this specification MUST NOT be sent to an SMTP server that does not support UTF8SMTP. Such a message MAY be rejected by a server if it lacks ALT-ADDRESSes as discussed in Section 3.2 of this specification.

The three-digit reply codes used in this section are consistent with their meanings as defined in RFC 2821.

When messages are rejected because the RCPT command requires an ALT-ADDRESS, the response code 553 is used with the meaning "mailbox name not allowed". When messages are rejected for other reasons, such as the MAIL command requiring an ALT-ADDRESS, the response code 550 is used with the meaning "mailbox unavailable". When the server supports enhanced mail system status codes [RFC3463], response code "X.6.7" [RFC5248] is used, 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 with the meaning "Transaction failed". When the server supports enhanced mail system status codes [RFC3463], response code "X.6.9" [RFC5248] is used, meaning that "UTF8SMTP downgrade failed".

3.6. Body Parts and SMTP Extensions

There is no ESMTP parameter to assert that a message is an internationalized message. An SMTP server that requires accurate knowledge of whether a message is internationalized is required to parse all message header fields and MIME header fields in the message body.

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, with the BODY=BINARYMIME parameter, if the server advertises BINARYMIME [RFC3030] and that is appropriate.

Assuming that the server advertises UTF8SMTP and 8BITMIME, and receives 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. If there is no BODY parameter, the header contains UTF-8 characters, but all the body parts are in ASCII (possibly as the result of a content-transfer-encoding).

2. If a BODY=8BITMIME parameter is present, the header contains UTF-8 characters, and some or all of the body parts contain 8-bit line-oriented data.

3. If a BODY=BINARYMIME parameter is present, the header contains UTF-8 characters, and some or all body parts contain binary data without restriction as to line lengths or delimiters.

3.7. Additional ESMTP Changes and Clarifications

The information carried in the mail transport process involves addresses ("mailboxes") and domain names in various 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 specification 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.

3.7.1. The Initial SMTP Exchange

When an SMTP connection is opened, the server normally sends a "greeting" response consisting of the 220 response 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.

3.7.2. Mail eXchangers

Organizations often 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 as described in RFC 2821. 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 users might perceive as a serious reliability issue.

3.7.3. 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" or "Received" 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, this memo updates 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 Section 4.4 of RFC 2821
    ; uReverse-path is defined in Section 3.3 of this document

uTime-stamp-line = "Received:" FWS uStamp <CRLF>
    ; Replaces Time-stamp-line in Section 4.4 of RFC 2821

uStamp = From-domain By-domain uOpt-info ";"  FWS date-time
    ; Replaces Stamp in Section 4.4 of RFC 2821
```

```
uOpt-info = [Via] [With] [ID] [uFor]
   ; Replaces Opt-info in Section 4.4 of RFC 2821
   ; The protocol value for With will allow a UTF8SMTP value

  uFor = "FOR" ( FWS (uPath / uMailbox) ) CFWS
   ; Replaces For in Section 4.4 of RFC 2821
   ; uPath and uMailbox are defined in Sections 2.4 and
   ; 2.3, respectively, of this document
```

Note: The FOR parameter has been changed to match the definition in [RFC2821bis], permitting only one address in the For clause. The group working on that document reached mailing list consensus that the syntax in [RFC2821] that permitted more than one address was simply a mistake.

Except in the 'uFor' clause and 'uReverse-path' value where non-ASCII domain names 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 when this extension is used is one of the UTF8SMTP values specified in the "IANA Considerations" section of this document.

3.7.4. UTF-8 Strings in Replies

3.7.4.1. MAIL and RCPT Commands

If the client issues a RCPT command containing non-ASCII characters, the SMTP server is permitted to use UTF-8 characters in the email address associated with 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 responses containing UTF-8 email addresses. If a given RCPT command does not include a non-ASCII envelope address, 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.

3.7.4.2. VRFY and EXPN Commands and the UTF8REPLY Parameter

If the VRFY and EXPN commands are transmitted with an optional parameter "UTF8REPLY", it indicates the client can accept UTF-8 strings in replies from those commands. This allows the server to use UTF-8 strings in mailbox names and full names that occur in replies without concern that the client might be confused by them. An SMTP client that conforms to this specification MUST accept and correctly process replies from the VRFY and EXPN commands that contain UTF-8 strings. However, the SMTP server MUST NOT use UTF-8

strings in replies if the SMTP client does not specifically allow such replies by transmitting this parameter. Most replies do not require that a mailbox name be included in the returned text, and therefore UTF-8 is not needed in them. Some replies, notably those resulting from successful execution of the VRFY and EXPN commands, do include the mailbox, making the provisions of this section important.

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

```
    "VRFY" SP (uLocal-part / uMailbox) [SP "UTF8REPLY"] CRLF
           ; uLocal-part and uMailbox are defined in
           ; Section 3.3 of this document.

    "EXPN" SP ( uLocal-part / uMailbox ) [ SP "UTF8REPLY" ] CRLF
           ; uLocal-part and uMailbox are defined in
           ; Section 3.3 of this document.
```

The "UTF8REPLY" parameter does not use a value. If the reply to a VERIFY (VRFY) or EXPAND (EXPN) command requires UTF-8, but the SMTP client does not use the "UTF8REPLY" parameter, then the server MUST use either the response code 252 or 550. Response code 252, defined in [RFC2821], means "Cannot VRFY user, but will accept the message and attempt the delivery". Response code 550, also defined in [RFC2821], means "Requested action not taken: mailbox unavailable". When the server supports enhanced mail system status codes [RFC3463], the enhanced response code as specified below is used. Using the "UTF8REPLY" parameter with a VERIFY (VRFY) or EXPAND (EXPN) command enables UTF-8 replies for that command only.

If a normal success response (i.e., 250) 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 3.3 of this document.
         ; User Name can contain non-ASCII characters.

      uMailbox
         ; uMailbox is defined in Section 3.3 of this document.
```

If the SMTP reply requires UTF-8 strings, but UTF-8 is not allowed in the reply, and the server supports enhanced mail system status codes [RFC3463], the enhanced response code is either "X.6.8" or "X.6.10" [RFC5248], meaning "A reply containing a UTF-8 string is required to show the mailbox name, but that form of response is not permitted by the client".

If the SMTP client does not support the UTF8SMTP extension, but receives a UTF-8 string in a reply, it may not be able to properly report the reply to the user, and some clients might crash. Internationalized messages in replies are only allowed in the commands under the situations described above. Under any other circumstances, UTF-8 text MUST NOT appear in the reply.

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 in replies except in the limited cases specifically permitted in this section.

4. IANA Considerations

IANA has added a new value "UTF8SMTP" to the SMTP Service Extension subregistry of the Mail Parameters registry, according to the following data:

Keywords	Description	Reference
UTF8SMTP	Internationalized email address	[RFC5336]

This document adds new values to the SMTP Enhanced Status Code subregistry of the Mail Parameters registry, following the guidance in Sections 3.5 and 3.7.4.2 of this document, and being based on [RFC5248]. The registration data is as follows:

```
Code:               X.6.7
Sample Text:        The ALT-ADDRESS is required but not specified
Associated basic status code:  553, 550
Description:        This indicates the reception of a MAIL or RCPT
                    command that required an ALT-ADDRESS parameter
                    but such parameter was not present.
Defined:            RFC 5336  (Experimental track)
Submitter:          Jiankang YAO
Change controller:  IESG.


Code:               X.6.8
Sample Text:        UTF-8 string reply is required,
                    but not permitted by the client
Associated basic status code:  553, 550
Description:        This indicates that a reply containing a UTF-8
                    string is required to show the mailbox name,
                    but that form of response is not
                    permitted by the client.
Defined:            RFC  5336.  (Experimental track)
Submitter:          Jiankang YAO
Change controller:  IESG.


 Code:               X.6.9
 Sample Text:        UTF8SMTP downgrade failed
 Associated basic status code:  550
 Description:        This indicates that transaction failed
                     after the final "." of the DATA command.
 Defined:            RFC  5336.  (Experimental track)
 Submitter:          Jiankang YAO
 Change controller:  IESG.


Code:               X.6.10
Sample Text:        UTF-8 string reply is required,
                    but not permitted by the client
Associated basic status code:  252
Description:        This indicates that a reply containing a UTF-8
                    string is required to show the mailbox name,
                    but that form of response is not
                    permitted by the client.
Defined:            RFC 5336.  (Experimental track)
Submitter:          Jiankang YAO
Change controller:  IESG.
```

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

WITH protocol types	Description	Reference
UTF8SMTP	UTF8SMTP with Service Extensions	[RFC5336]
UTF8SMTPA	UTF8SMTP with SMTP AUTH	[RFC4954] [RFC5336]
UTF8SMTPS	UTF8SMTP with STARTTLS	[RFC3207] [RFC5336]
UTF8SMTPSA	UTF8SMTP with both STARTTLS and SMTP AUTH	[RFC3207] [RFC4954] [RFC5336]

5. Security Considerations

See the extended security considerations discussion in the framework document [RFC4952].

6. Acknowledgements

Much of the text in the initial version of this specification was derived or copied from [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 specification. Additional important comments and suggestions, and often specific text, were contributed by many members of the WG and design team. Those contributions include material from 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, Pete Resnick, S. Moonesamy, Soobok Lee, Shawn Steele, Alfred Hoenes, Miguel Garcia, Magnus Westerlund, and Lars Eggert. Of course, none of the individuals are necessarily responsible for the combination of ideas represented here.

7. References

7.1. Normative References

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

[RFC1652] Klensin, J., Freed, N., Rose, M., Stefferud, E., and D. Crocker, "SMTP Service Extension for 8bit-MIMEtransport", RFC 1652, July 1994.

[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", STD 63, RFC 3629, November 2003.

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

[RFC4952] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 4952, July 2007.

[RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.

[RFC5248] Hansen, T. and J. Klensin, "A Registry for SMTP Enhanced Mail System Status Codes", BCP 138, RFC 5248, June 2008.

[RFC5335] Abel, Y., Ed., "Internationalized Email Headers", RFC 5335, September 2008.

[RFC5337] Newman, C. and A. Melnikov, Ed., "Internationalized Delivery Status and Disposition Notifications", RFC 5337, September 2008.

7.2. Informative References

[Downgrade] Fujiwara, K. and Y. Yoneya, "Downgrading mechanism for Email Address Internationalization", Work in Progress, July 2008.

[Emailaddr] Klensin, J., "Internationalization of Email Addresses", Work in Progress, July 2005.

[RFC0974] Partridge, C., "Mail routing and the domain system", RFC 974, January 1986.

[RFC2033] Myers, J., "Local Mail Transfer Protocol", RFC 2033, October 1996.

[RFC2821bis] Klensin, J., "Simple Mail Transfer Protocol", Work in Progress, July 2008.

[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.

[RFC4954] Siemborski, R., Ed. and A. Melnikov, Ed., "SMTP Service Extension for Authentication", RFC 4954, July 2007.

Appendix A. Material Updating RFC 4952

RFC 4952, the overview and framework document covering this set of extensions for internationalized email, was completed before this specification, which specifies a particular part of the protocol set. This appendix, which is normative, contains material that would have been incorporated into RFC 4952 had it been delayed until the work described in the rest of this specification was completed. This material should be included in any update to RFC 4952.

A.1. Conventional Message and Internationalized Message

- A conventional message is one that does not use any extension defined in this document or in the UTF-8 header specification [RFC5335], and which is strictly conformant to RFC 2822 [RFC2822].

- An internationalized message is a message utilizing one or more of the extensions defined in this specification or in the UTF-8 header specification [RFC5335], so that it is no longer conformant to the RFC 2822 specification of a message.

A.2. 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 needs to be updated to deal with these situations.

A.3. SMTP Service Extension for DSNs

The existing Draft Standard regarding delivery status notifications (DSNs) [RFC3461] is limited to ASCII text in the machine readable portions of the protocol. "International Delivery Status and Disposition Notifications" [RFC5337] adds a new address type for international email addresses so an original recipient address with non-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 [RFC5337] including support for the ORCPT parameter.

A.4. 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.

A.5. Applicability of SMTP Extension to Additional Uses

Among other protocol changes, the SMTP extension allows an optional alternate address to be supplied with the MAIL and RCPT commands. For the purposes of this set of specifications, this alternate address only has meaning when the primary address contains UTF-8 characters and the message is downgraded. While it may be tempting to consider the alternate address as a general-purpose second-chance address to be used whenever the primary address is rejected, such behavior is not defined here. This restriction allows for future extensions to be developed which create such a general-purpose second-chance address, although no specific work on such an extension is currently anticipated. Note that any such extension needs to consider the question of what the [RFC0974] sequencing rules mean when different possible servers support different sets of ESMTP options (or, in this case, addresses). The answer to this question may also imply updates to [RFC2821].

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 58812230
EMail: maowei_ietf@cnnic.cn

Full Copyright Statement

Intellectual Property