Network Working Group
Internet-Draft
Obsoletes: RFC5721
(if approved)
Updates: RFC1939
(if approved)
Intended status: Standards Track
Expires: March 14, 2011

R. Gellens
QUALCOMM Incorporated
C. Newman
Oracle
Jiankang. Yao
CNNIC
Kazunori. Fujiwara
JPRS

September 28, 2010

POP3 Support for UTF-8
draft-ietf-eai-rfc5721bis-00.txt

Abstract

This specification extends the Post Office Protocol version 3 (POP3) to support un-encoded international characters in user names, passwords, mail addresses, message headers, and protocol-level textual error strings.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on March 14, 2011.

Copyright Notice

Table of Contents

1. Introduction 3
  1.1. Conventions Used in This Document 3
2. LANG Capability 4
3. UTF8 Capability 6
  3.1. The UTF8 Command 7
  3.2. USER Argument to UTF8 Capability 8
4. Native UTF-8 Maildrops 9
5. IANA Considerations 9
6. Security Considerations 9
7. References 10
  7.1. Normative References 10
  7.2. Informative References 11
Appendix A. Design Rationale 12
Appendix B. Acknowledgments 12

1. Introduction

This document forms part of the Email Address Internationalization (EAI) protocols described in the EAI Framework document [I-D.ietf-eai-frmwrk-4952bis]. As part of the overall EAI work, email messages may be transmitted and delivered containing un-encoded UTF-8 characters, and mail drops that are accessed using POP3 [RFC1939] might natively store UTF-8.

This specification extends POP3 [RFC1939] using the POP3 extension mechanism [RFC2449] to permit un-encoded UTF-8 [RFC3629] in headers, as described in "Internationalized Email Headers" [I-D.ietf-eai-rfc5335bis]. It also adds a mechanism to support login names and passwords outside the ASCII character set, and a mechanism to support UTF-8 protocol-level error strings in a language appropriate for the user.

Within this specification, the term "down-conversion" refers to the process of modifying a message containing UTF-8 headers [I-D.ietf-eai-rfc5335bis] or body parts with 8bit content-transfer-encoding, as defined in MIME Section 2.8 [RFC2045], into conforming 7-bit Internet Message Format [RFC5322] with message header extensions for non-ASCII text [RFC2047] and other 7-bit encodings. Down-conversion is specified by "Message-Downgrading for Email Address Internationalization" [message-downgrade].

1.1. Conventions Used in This Document

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 "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

The formal syntax uses the Augmented Backus-Naur Form (ABNF) [RFC5234] notation, including the core rules defined in Appendix B of RFC 5234.

In examples, "C:" and "S:" indicate lines sent by the client and server, respectively. If a single "C:" or "S:" label applies to multiple lines, then the line breaks between those lines are for editorial clarity only and are not part of the actual protocol exchange.

Note that examples always use 7-bit ASCII characters due to limitations of this document format; in particular, some examples for the "LANG" command may appear silly as a result.

2. LANG Capability

Per "POP3 Extension Mechanism" [RFC2449], this document adds a new capability response tag to indicate support for a new command: LANG. The capability tag and new command are described below.

CAPA tag:
LANG

Arguments with CAPA tag:
none

Added Commands:
LANG

Standard commands affected:
All

Announced states / possible differences:
both / no

Commands valid in states:
AUTHENTICATION, TRANSACTION

Specification reference:
this document

Discussion:

POP3 allows most +OK and -ERR server responses to include human-readable text that, in some cases, might be presented to the user. But that text is limited to ASCII by the POP3 specification [RFC1939]. The LANG capability and command permit a POP3 client to negotiate which language the server should use when sending human-readable text.

A server that advertises the LANG extension MUST use the language "i-default" as described in [RFC2277] as its default language until another supported language is negotiated by the client. A server MUST include "i-default" as one of its supported languages.

The LANG command requests that human-readable text included in all subsequent +OK and -ERR responses be localized to a language matching the language range argument (the "Basic Language Range" as described by [RFC4647]). If the command succeeds, the server returns a +OK response followed by a single space, the exact language tag selected, another space, and the rest of the line is human-readable text in the appropriate language. This and subsequent protocol-level human-

readable text is encoded in the UTF-8 charset.

If the command fails, the server returns an -ERR response and subsequent human-readable response text continues to use the language that was previously active (typically i-default).

The special "*" language range argument indicates a request to use a language designated as preferred by the server administrator. The preferred language MAY vary based on the currently active user.

If no argument is given and the POP3 server issues a positive response, then the response given is multi-line. After the initial +OK, for each language tag the server supports, the POP3 server responds with a line for that language. This line is called a "language listing".

In order to simplify parsing, all POP3 servers are required to use a certain format for language listings. A language listing consists of the language tag [RFC5646] of the message, optionally followed by a single space and a human-readable description of the language in the language itself, using the UTF-8 charset.

Examples:

```
< Note that some examples do not include the correct character
accents due to limitations of this document format. >

< The server defaults to using English i-default responses until
the client explicitly changes the language. >

C: USER karen
S: +OK Hello, karen
C: PASS password
S: +OK karen's maildrop contains 2 messages (320 octets)

< Client requests deprecated MUL language.  Server replies
with -ERR response. >

C: LANG MUL
S: -ERR invalid language MUL

< A LANG command with no parameters is a request for
a language listing. >

C: LANG
S: +OK Language listing follows:
S: en English
S: en-boont English Boontling dialect
```

```
S: de Deutsch
S: it Italiano
S: es Espanol
S: sv Svenska
S: i-default Default language
S: .

< A request for a language listing might fail. >

C: LANG
S: -ERR Server is unable to list languages

< Once the client changes the language, all responses will be in
that language, starting with the response to the LANG command. >

C: LANG es
S: +OK es Idioma cambiado

< If a server does not support the requested primary language,
responses will continue to be returned in the current language
the server is using. >

C: LANG uga
S: -ERR es Idioma <<UGA>> no es conocido

C: LANG sv
S: +OK sv Kommandot "LANG" lyckades

C: LANG *
S: +OK es Idioma cambiado
```

3. UTF8 Capability

Per "POP3 Extension Mechanism" [RFC2449], this document adds a new capability response tag to indicate support for new server functionality, including a new command: UTF8. The capability tag and new command and functionality are described below.

CAPA tag:
UTF8

Arguments with CAPA tag:
USER

Added Commands:
UTF8

Standard commands affected:
USER, PASS, APOP, LIST, TOP, RETR

Announced states / possible differences:
both / no

Commands valid in states:
AUTHORIZATION

Specification reference:
this document

Discussion:

This capability adds the "UTF8" command to POP3. The UTF8 command switches the session from ASCII to UTF-8 mode.

3.1. The UTF8 Command

The UTF8 command enables UTF-8 mode. The UTF8 command has no parameters.

Maildrops can natively store UTF-8 or be limited to ASCII. UTF-8 mode has no effect on messages in an ASCII-only maildrop. Messages in native UTF-8 maildrops can be ASCII or UTF-8 using internationalized headers [I-D.ietf-eai-rfc5335bis] and/or 8bit content-transfer-encoding, as defined in MIME Section 2.8 [RFC2045]. In UTF-8 mode, both UTF-8 and ASCII messages are sent to the client as-is (without conversion). When not in UTF-8 mode, UTF-8 messages in a native UTF-8 maildrop MUST NOT be sent to the client as-is. UTF-8 messages in a native UTF-8 maildrop MUST be down-converted (downgraded) to comply with unextended POP and Internet Mail Format without UTF-8 mode support.

Note that even in UTF-8 mode, MIME binary content-transfer-encoding is still not permitted.

The octet count (size) of a message reported in a response to the LIST command SHOULD match the actual number of octets sent in a RETR response (not counting byte-stuffing). Sizes reported elsewhere, such as in STAT responses and non-standardized, free-form text in positive status indicators (following "+OK") need not be accurate, but it is preferable if they are.

Mail stores are either ASCII or native UTF-8, and clients either issue the UTF8 command or not. The message needs converting only when it is native UTF-8 and the client has not issued the UTF8 command, in which case the server must down-convert it. The down-

converted message may be larger. The server may choose various strategies regarding down-conversion, which include when to down-convert, whether to cache or store the down-converted form of a message (and if so, for how long), and whether to calculate or retain the size of a down-converted message independently of the down-converted content. If the server does not have immediate access to the accurate down-converted size, it may be faster to estimate rather than calculate it. Servers are expected to normally follow the RFC 1939 [RFC1939] text on using the "exact size" in a scan listing, but there may be situations with maildrops containing very large numbers of messages in which this might be a problem. If the server does estimate, reporting a scan listing size smaller than what it turns out to be could be a problem for some clients. In summary, it is better for servers to report accurate sizes, but if this is not possible, high guesses are better than small ones. Some POP servers include the message size in the non-standardized text response following "+OK" (the 'text' production of RFC 2449 [RFC2449]), in a RETR or TOP response (possibly because some examples in POP3 [RFC1939] do so). There has been at least one known case of a client relying on this to know when it had received all of the message rather than following the POP3 [RFC1939] rule of looking for a line consisting of a termination octet (".") and a CRLF pair. While any such client is non-compliant, if a server does include the size in such text, it is better if it is accurate.

Clients MUST NOT issue the STLS command [RFC2595] after issuing UTF8; servers MAY (but are not required to) enforce this by rejecting with an "-ERR" response an STLS command issued subsequent to a successful UTF8 command. (Because this is a protocol error as opposed to a failure based on conditions, an extended response code [RFC2449] is not specified.)

3.2. USER Argument to UTF8 Capability

If the USER argument is included with this capability, it indicates that the server accepts UTF-8 user names and passwords.

Servers that include the USER argument in the UTF8 capability response SHOULD apply SASLprep [RFC4013] to the arguments of the USER and PASS commands.

A client or server that supports APOP and permits UTF-8 in user names or passwords MUST apply SASLprep [RFC4013] to the user name and password used to compute the APOP digest.

When applying SASLprep [RFC4013], servers MUST reject UTF-8 user names or passwords that contain a Unicode character listed in Section 2.3 of SASLprep [RFC4013]. When applying SASLprep to the USER

argument, the PASS argument, or the APOP username argument, a compliant server or client MUST treat them as a query string (i.e., unassigned Unicode code points are allowed). When applying SASLprep to the APOP password argument, a compliant server or client MUST treat them as a stored string (i.e., unassigned Unicode code points are prohibited).

The client does not need to issue the UTF8 command prior to using UTF-8 in authentication. However, clients MUST NOT use UTF-8 characters in USER, PASS, or APOP commands unless the USER argument is included in the UTF8 capability response.

The server MUST reject UTF-8 user names or passwords that fail to comply with the formal syntax in UTF-8 [RFC3629].

Use of UTF-8 characters in the AUTH command is governed by the POP3 SASL [RFC5034] mechanism.

4. Native UTF-8 Maildrops

When a POP3 server uses a native UTF-8 maildrop, it is the responsibility of the server to comply with the POP3 base specification [RFC1939] and Internet Message Format [RFC5322] when not in UTF-8 mode. Mechanisms for 7-bit downgrading to help comply with the standards are described in [message-downgrade].

5. IANA Considerations

This specification adds two new capabilities ("UTF8" and "LANG") to the POP3 capability registry [RFC2449].

6. Security Considerations

The security considerations of UTF-8 [RFC3629] and SASLprep [RFC4013] apply to this specification, particularly with respect to use of UTF-8 in user names and passwords.

The "LANG *" command might reveal the existence and preferred language of a user to an active attacker probing the system if the active language changes in response to the USER, PASS, or APOP commands prior to validating the user's credentials. Servers MUST implement a configuration to prevent this exposure.

It is possible for a man-in-the-middle attacker to insert a LANG command in the command stream, thus making protocol-level diagnostic responses unintelligible to the user. A mechanism to integrity-protect the session, such as Transport Layer Security (TLS) [RFC2595] can be used to defeat such attacks.

Modifying server authentication code (in this case, to support UTF8 command) needs to be done with care to avoid introducing vulnerabilities (for example, in string parsing).

The UTF8 command description (Section 3.1) contains a discussion on reporting inaccurate sizes. An additional risk to doing so is that, if a client allocates buffers based on the reported size, it may overrun the buffer, crash, or have other problems if the message data is larger than reported.

7. References

7.1. Normative References

[I-D.ietf-eai-frmwrk-4952bis] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", draft-ietf-eai-frmwrk-4952bis-07 (work in progress), August 2010.

[I-D.ietf-eai-rfc5335bis] Yang, A. and S. Steele, "Internationalized Email Headers", draft-ietf-eai-rfc5335bis-02 (work in progress), August 2010.

[RFC1939] Myers, J. and M. Rose, "Post Office Protocol - Version 3", STD 53, RFC 1939, May 1996.

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

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

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.

[RFC2449] Gellens, R., Newman, C., and L. Lundblade, "POP3 Extension Mechanism", RFC 2449, November 1998.

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

[RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", RFC 4013, February 2005.

[RFC4647] Phillips, A. and M. Davis, "Matching of Language Tags", BCP 47, RFC 4647, September 2006.

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

[RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.

[RFC5646] Phillips, A. and M. Davis, "Tags for Identifying Languages", BCP 47, RFC 5646, September 2009.

7.2. Informative References

[RFC2595] Newman, C., "Using TLS with IMAP, POP3 and ACAP", RFC 2595, June 1999.

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

[RFC5034] Siemborski, R. and A. Menon-Sen, "The Post Office Protocol (POP3) Simple Authentication and Security Layer (SASL) Authentication Mechanism", RFC 5034, July 2007.

[message-downgrade] Fujiwara, K. and Y. Yoneya, "Message Downgrading for Email Address Internationalization (EAI) Maildrops", draft-ietf-eai-rfc5504bis-00 (work in progress), Sep 2010.

Appendix A. Design Rationale

This non-normative section discusses the reasons behind some of the design choices in the above specification.

Due to interoperability problems with RFC 2047 and limited deployment of RFC 2231, it is hoped these 7-bit encoding mechanisms can be deprecated in the future when UTF-8 header support becomes prevalent.

USER is optional because the implementation burden of SASLprep [RFC4013] is not well understood, and mandating such support in all cases could negatively impact deployment.

While it is possible to provide useful examples for language negotiation without support for non-ASCII characters, it is difficult to provide useful examples for commands specifically designed to use the UTF-8 charset un-encoded when the document format is limited to ASCII. As a result, there are no plans to provide examples for that part of the specification as long as this remains an experimental proposal. However, implementers of this specification are encouraged to provide examples to the document authors for a future revision.

Appendix B. Acknowledgments

Thanks to John Klensin, Tony Hansen, and other EAI working group participants who provided helpful suggestions and interesting debate that improved this specification.

Authors' Addresses

Randall Gellens
QUALCOMM Incorporated
5775 Morehouse Drive
San Diego, CA 92651
US

EMail: rg+ietf@qualcomm.com

Chris Newman
Oracle
800 Royal Oaks
Monrovia, CA 91016-6347
US

EMail: chris.newman@oracle.com

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

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

Kazunori Fujiwara
Japan Registry Services Co., Ltd.
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Tokyo

Phone: +81 3 5215 8451
EMail: fujiwara@jprs.co.jp