POP3 Support for UTF-8
draft-ietf-eai-pop-03.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 27, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2008).

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.
Table of Contents

1. Introduction .................................................. 3
   1.1. Conventions Used in this Document ....................... 3
   1.2. Change History ............................................ 3
      1.2.1. Changes from -02 to -03 .......................... 3
      1.2.2. Changes from -01 to -02 ......................... 4
      1.2.3. Changes from -00 to -01 ......................... 4
      1.2.4. Changes from draft-newman-ima-pop .................. 4
   1.3. Open Issues ................................................ 5
2. LANG Capability ................................................. 5
3. UTF8 Capability ................................................ 8
   3.1. The UTF8 Command .......................................... 8
   3.2. USER Argument to UTF8 Capability ....................... 9
4. Issues with UTF-8 Header maildrop ............................. 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 .................................... 11
Appendix B. Acknowledgments ..................................... 12
Authors’ Addresses .............................................. 12
Intellectual Property and Copyright Statements .................. 13
1. Introduction

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-utf8headers]. It also adds a mechanism to support login names 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 UTF8 headers [I-D.ietf-eai-utf8headers] or body parts with 8bit content-transfer-encoding as defined in MIME section 2.8 [RFC2045] into conforming 7-bit Internet message format [RFC2822] with Message Header Extensions for Non-ASCII Text [RFC2047] and other 7-bit encodings. Down-conversion is specified by Downgrading mechanism for Email Address Internationalization [I-D.ietf-eai-downgrade].

1.1. Conventions Used in this Document

The key words "MUST", "MUST NOT", "SHOULD", "SHOULD NOT", and "MAY" in this document are to be interpreted as defined in "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

The formal syntax uses the Augmented Backus-Naur Form (ABNF) [RFC4234] notation including the core rules defined in Appendix B of RFC 4234.

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.

1.2. Change History

This section describes the change history of this Internet draft and will be removed when/if this is published as an RFC.

1.2.1. Changes from -02 to -03

- Updated references.
- Replaced US-ASCII with ASCII.
- Added comment to language listing failure example.
Replaced RET8, LST8, and TOP8 commands with a single mode-switch UTF8 command issued before authentication. This simplifies the protocol, and allows servers to optionally down-convert a cache of the maildrop prior to issuing the +OK response entering TRANSACTION state.

Removed most up-conversion material.

Removed definition of up-conversion.

Removed IMAP4 reference.

Added AUTH command to those affected by UTF8 capability.

Removed LST8 and TOP8 capability parameters and commands.

Removed NO-RETR capability. POP servers are now unconditionally required to support down-conversion of UTF8-native maildrops.

Added sentence about modifying authentication code to Security Considerations.

eai-downgrade draft is now normative and required.

Deleted references to RFCs 1341, 1847, 2049, 2183, 3501, 3516, and 3490.

1.2.2. Changes from -01 to -02

- Minor grammatical tweaks.
- Add passwords to Abstract.
- Removed new editor’s name from Acknowledgments.

1.2.3. Changes from -00 to -01

- Update references

1.2.4. Changes from draft-newman-ima-pop

- Change title to make this a WG document.
- Add LANG command and extension.
- Rename RET8 capability to UTF8 and add sub-sections for arguments.
o Add TOP8 command.

o Add definition of up-conversion and down-conversion.

o Some grammar fix-ups and section re-ordering based on RFC editor style.

1.3. Open Issues

1. POP servers are now unconditionally required to down-convert messages in a UTF8-native maildrop when not in UTF8 mode. Previous versions of this document permitted servers to decline to do so. This should be verified as group consensus.

2. Servers do not up-convert messages. While this makes sense for RETR (since doing so would break digital signatures), it might make sense for TOP, as it would provide a smaller and more clear representation of the message. However, the usefulness of this seems to be not worth the extra burden.

3. The current version allows clients to use UTF8 in authentication commands without sending the UTF8 command.

2. LANG Capability

CAPA tag:

LANG

Arguments:

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 needs to 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 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 [RFC4646] of the message, optionally followed by a single space and a human readable description of that language using the UTF-8 charset.

< 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 depricated MUL language. Server replies with -ERR response.)

C: LANG MUL
S: -ERR invalid language MUL

(A LANG command with no arguments 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 German
S: it Italian
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. Note: the example does not include the correct character accents due to limitations of this document format.)

C: LANG fr
S: +OK fr La Language commande a ete execute avec success

(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 Ce Language n’est pas supporte

C: LANG fr-ca
S: +OK fr La Language commande a ete execute avec success

C: LANG *
S: +OK fr La Language commande a ete execute avec success

Examples
3. UTF8 Capability

CAPA tag:
UTF8

Arguments:
USER, LIST, TOP

Added Commands:
UTF8

Standard commands affected:
AUTH, 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 UTF8 mode.

3.1. The UTF8 Command

The UTF8 command enables UTF8 mode. Maildrops can natively store UTF8 or be limited to ASCII. UTF8 mode has no effect on messages in an ASCII-only maildrop. Messages in native-UTF8 maildrops can be ASCII or UTF8 using internationalized headers [I-D.ietf-eai-utf8headers] and/or 8bit content-transfer-encoding as defined in MIME section 2.8 [RFC2045]. In UTF8 mode, both UTF8 and ASCII messages are sent to the client as-is (without conversion). When not in UTF8 mode, UTF8 messages in a native UTF8 maildrop MUST be downconverted using the Downgrading mechanism for Email Address Internationalization [I-D.ietf-eai-downgrade].

Note that even in UTF8 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 what the server sends in a RETR response. Sizes reported elsewhere, such as in STAT responses and free-form text in positive status indicators (following "+OK") need not be
accurate, but it is preferable if they are.

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 and applies SASLprep [RFC4013] to the arguments of the AUTH, USER, PASS and APOP commands. A client that supports APOP and permits UTF-8 in user names or passwords MUST also implement SASLprep [RFC4013] on the user name and password used to compute the APOP digest.

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

Use of UTF8 in the AUTH command is governed by the SASL mechanism.

4. Issues with UTF-8 Header maildrop

When a POP3 server uses a UTF8-native maildrop, it is the responsibility of the server to comply with the POP3 base specification [RFC1939] and RFC 2822 [RFC2822] when not in UTF8 mode. Mechanisms for 7-bit downgrading to help comply with the standards are specified in Downgrading mechanism for Email Address Internationalization [I-D.ietf-eai-downgrade].

5. IANA Considerations

This 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 can 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 TLS [RFC2595] can be used to defeat such attacks.

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

7. References

7.1. Normative References


7.2. Informative References


Appendix A. Design Rationale

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

Having servers perform up-conversion so that, at a minimum, RFC2047-encoded words are decoded into UTF8 is tempting, since this is an area that clients often fail to correctly implement. However, modifying messages breaks digital signatures, and would require servers to support arbitrary charset conversion.

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.

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.

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 author 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

Chris Newman  
Sun Microsystems  
3401 Centrelake Dr., Suite 410  
Ontario, CA  91761  
US  
Email: chris.newman@sun.com

Randall Gellens  
QUALCOMM Incorporated  
5775 Morehouse Drive  
San Diego, CA  92651  
US  
Email: rg+ietf@qualcomm.com
Full Copyright Statement

Copyright (C) The IETF Trust (2008).

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.

Acknowledgment

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