Network Working Group   C. Newman
Internet-Draft   Sun Microsystems
Updates: 1939 (if approved)   February 27, 2006
Expires: August 31, 2006

                        POP3 Support for UTF-8
                      draft-newman-ima-pop-00.txt

Status of this Memo

Copyright Notice

Abstract

   This specification extends the Post Office Protocol version 3 (POP3)
   to support unencoded international characters in user names, mail
   addresses and message headers. This is an early draft and intended
   as a framework for discussion. Please do not deploy implementations
   of this draft.

Table of Contents

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

2. Introduction

This specification extends POP3 [RFC1939] using the POP3 Extension
Mechanism [RFC2449] to permit unencoded UTF-8 [RFC3629] in headers as
described in Transmission of Email Headers in UTF-8 Encoding
[I-D.yeh-ima-utf8headers]. It also adds a mechanism to support login
names outside the US-ASCII character set.

3. RET8 Capability

CAPA tag:
   UTF8

Arguments:
   USER, LST8

Added Commands:
   RET8, LST8

Standard commands affected:
   USER, PASS, APOP

Announced states / possible differences:
   both / no

Commands valid in states:
   TRANSACTION

Specification reference:
   this document

Discussion:

This capability adds UTF-8 support to POP3. This capability always
adds the "RET8" command to POP3. The RET8 command is identical to
the RETR command, except that the retrieved message uses UTF-8 in
headers [I-D.yeh-ima-utf8headers]. In addition, the 8bit content-
transfer-encoding as defined in MIME section 2.8 [RFC2045] is
explicitly permitted. The retrieved message MUST still be textual

and otherwise formatted according to RFC 2822 [RFC2822] and MIME
[RFC2045].   The MIME binary content-transfer-encoding is not
permitted. Clients wishing to use binary MIME should implement IMAP4
[RFC3501] with the IMAP4 Binary Content Extension [RFC3516].

If the USER argument is included with this capability, that indicates
the server accepts UTF-8 user names and passwords and applies
SASLprep [RFC4013] to the arguments of the USER, PASS and APOP
commands. A client which 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.

If the LST8 argument is included with this capability, that indicates
the server implements the LST8 command. The LST8 command is
identical to the LIST command except that the octet counts are the
exact octet counts returned by the RET8 command. A POP3 client which
uses RET8 MUST use LST8 instead of LIST if LST8 is advertised.

4. NO-RETR Capability

CAPA tag:
   NO-RETR

Arguments:
   none

Added Commands:
   none

Standard commands affected:
   RETR, LIST, TOP

Announced states / possible differences:
   both / no

Commands valid in states:
   N/A

Specification reference:
   this document

Discussion:

This capability permits a POP3 server to advertise that it does not
support the RETR, LIST or TOP commands. Any attempt to use any of
these three commands will result in an error response. As this is an
incompatible change to POP3, a clear warning is necessary. POP3
clients which find implementation of the UTF8 capability problematic

   are encouraged to at least detect the NO-RETR capability and provide
   an informative error message to the end-user.

   When a POP3 server runs on a UTF-8 header native mail drop, the down-
   conversion step necessary to implement RETR in a backwards compatible
   fashion will become more difficult to support. Although it is hoped
   deployed POP3 servers do not advertise NO-RETR for some years, this
   capability is intended to minimize the disruption when legacy support
   finally goes away.

   A server which advertises NO-RETR MUST advertise UTF8 with at least
   the LST8 argument and MUST NOT advertise TOP.

5. Up-Conversion Server Requirements

   When a POP3 server uses a traditional mail drop that supports only
   7-bit headers, it MUST support message header up-conversion for the
   RET8 and LST8 commands. As POP3 clients are best when simple, the
   more up-conversion the server performs, the better. Minimal up-
   conversion is described in this section.

   The server MUST support up-conversion of the following address
   header-fields in the message header: From, Sender, To, CC, Bcc,
   Resent-From, Resent-Sender, Resent-To, Resent-CC, Resent-Bcc, and
   Reply-To. This up-conversion MUST include address local-parts
   encoded according to [TBD], address domains encoded according to IDNA
   [RFC3490], and MIME header encoding [RFC2047] of display-names and
   any RFC 2822 comments.

   The following charsets MUST be supported for up-conversion of MIME
   header encoding [RFC2047]: UTF-8, US-ASCII, ISO-8859-1, ISO-8859-2,
   ISO-8859-3, ISO-8859-4, ISO-8859-5, ISO-8859-6, ISO-8859-7,
   ISO-8859-8, ISO-8859-9, ISO-8859-10, ISO-8859-14, and ISO-8859-15.
   Other widely deployed MIME charsets SHOULD be supported.

   Up-conversion of MIME header encoding of the following headers MUST
   also be implemented: Subject, Date (RFC 2822 comments only),
   Comments, Keywords, Content-Description.

   While this specification does not require it, server implementations
   are encouraged to up-convert all MIME body headers, and particularly
   the deprecated (and misused) name parameter [RFC1341] on Content-Type
   and the Content-Disposition filename parameter. These may be encoded
   using the standard MIME parameter encoding [RFC2231] mechanism, or
   via non-standard use of MIME header encoding [RFC2047] in quoted
   strings.

   The POP server MUST NOT perform up-conversion of headers and content

of multipart/signed, as well as Original-Recipient and Return-Path.

6. Issues with UTF-8 Header Mail Drop

When a POP3 server uses a mail drop that supports UTF-8 headers and
it does not advertise the NO-RETR capability, it is the
responsibility of the server to comply with the POP3 base
specification [RFC1939] and RFC 2822 [RFC2822] with respect to the
RETR, LIST and TOP commands. Mechanisms for 7-bit downgrading to
help comply with the standards are discussed in Downgrading mechanism
for Internationalized eMail Address (IMA) [I-D.yoneya-ima-downgrade].

A POP3 server with a mail drop that supports UTF-8 headers MUST
comply with the RET8 protocol requirements implicit from Section 5.
However, the code necessary for such compliance need not be part of
the POP3 server itself in this case. For example, the minimal
required up-conversion could be performed when a message is inserted
into the POP3-accessible mail drop.

7. IANA Considerations

This adds two new capabilities ("UTF8" and "NO-RETR") to the POP3
capability registry [RFC2449].

8. 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 server. Otherwise, this is not
believed to alter the security considerations of POP3.

9. References

9.1 Normative References

[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.

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

   [RFC2822] Resnick, P., "Internet Message Format", RFC 2822,
              April 2001.

   [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.

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

   [RFC4234] Crocker, D. and P. Overell, "Augmented BNF for Syntax
              Specifications: ABNF", RFC 4234, October 2005.

   [I-D.yeh-ima-utf8headers]
              Yeh, J., "Transmission of Email Headers in UTF-8
              Encoding", draft-yeh-ima-utf8headers-00 (work in
              progress), September 2005.

9.2 Informative References

   [RFC1341] Borenstein, N. and N. Freed, "MIME (Multipurpose Internet
              Mail Extensions): Mechanisms for Specifying and Describing
              the Format of Internet Message Bodies", RFC 1341,
              June 1992.

   [RFC1847] Galvin, J., Murphy, S., Crocker, S., and N. Freed,
              "Security Multiparts for MIME: Multipart/Signed and
              Multipart/Encrypted", RFC 1847, October 1995.

   [RFC2049] Freed, N. and N. Borenstein, "Multipurpose Internet Mail
              Extensions (MIME) Part Five: Conformance Criteria and
              Examples", RFC 2049, November 1996.

   [RFC2183] Troost, R., Dorner, S., and K. Moore, "Communicating
              Presentation Information in Internet Messages: The
              Content-Disposition Header Field", RFC 2183, August 1997.

   [RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded
              Word Extensions: Character Sets, Languages, and
              Continuations", RFC 2231, November 1997.

   [RFC2277] Alvestrand, H., "IETF Policy on Character Sets and

                Languages", BCP 18, RFC 2277, January 1998.

   [RFC3501]   Crispin, M., "INTERNET MESSAGE ACCESS PROTOCOL - VERSION
               4rev1", RFC 3501, March 2003.

   [RFC3516]   Nerenberg, L., "IMAP4 Binary Content Extension", RFC 3516,
               April 2003.

   [I-D.yoneya-ima-downgrade]
               Yoneya, Y. and K. Fujiwara, "Downgrading mechanism for
               Internationalized eMail Address (IMA)",
               draft-yoneya-ima-downgrade-00 (work in progress),
               October 2005.

Author's Address

   Chris Newman
   Sun Microsystems
   3401 Centrelake Dr., Suite 410
   Ontario, CA 91761
   US

   Email: chris.newman@sun.com

Appendix A. Design Rationale

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

   The basic approach of advertising a parallel command set and
   permitting graceful migration of both client and server with minimal
   disruption is a deliberate choice. While a mechanism that makes RETR
   "just-send-UTF-8" might deploy faster, it would also create
   interoperability problems. The approach used prevents
   interoperability problems until the NO-RETR mechanism is deployed. A
   client command to cause a model switch could also work, but the
   parallel command approach is cleaner given the small number of
   commands.

   The choice to make RET8 nearly identical to RETR is important to
   minimize the code changes necessary in a client. An alternative
   approach which permits binary MIME and uses a length-counted argument
   would be architecturally superior but is dismissed due to the
   migration problems it would cause. The IMAP4 Binary extension should
   be sufficient for cases where binary MIME support is deemed
   necessary.

LST8 is optional to minimize the cost of deploying UTF-8 support on a
legacy mail drop. The server load necessary to perform up-conversion
on every message in the mail drop to determine the LST8 octet-counts
would be prohibitively expensive when there's no way to cache those
counts. The octet counts from the LIST command should be close
enough to the RET8 size for most POP3 user interfaces, and robust
POP3 clients already have to deal with LIST octet counts that don't
match the actual size of the RETR result.

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.

The NO-RETR mechanism simplifies diagnosis of interoperability
problems when legacy support goes away. In the situation where
backwards compatibility is broken anyway, just-send-8 RETR has the
advantage that it might work with some legacy clients. However, the
difficulty of diagnosing interoperability problems caused by a just-
send-8 RETR mechanism is the reason the NO-RETR mechanism was chosen.

This specification deliberately deprecates the optional TOP command
by not providing a TOP8 command. TOP is a crude partial fetch
mechanism, especially now that MIME support is widespread. IMAP4rev1
[RFC3501] now has complete partial fetch functionality. As a result
it is preferable to error on the side of simplicity in this case.

The up-conversion requirements are designed to balance the desire to
deprecate and eventually eliminate complicated encodings (like MIME
header encodings) without creating a significant deployment burden
for servers. While it would be desirable to require up-conversion of
attachment file names, the erroneous perception that MIME parsing is
difficult in combination with multiple deployed mechanisms for such
file names tip the balance.

The set of mandatory charsets comes from two sources: MIME
requirements [RFC2049] and IETF Policy on Character Sets [RFC2277].
Including a requirement to up-convert widely deployed encoded
ideographic charsets to UTF-8 would be reasonable for most scenarios,
but may require unacceptable table sizes for some embedded devices.
The open-ended recommendation to support widely deployed charsets
avoids the political ramifications of attempting to list such
charsets. The author believes market forces, existing open-source
software, and public conversion tables are sufficient to deploy the
appropriate charsets.

Appendix B. Acknowledgments

TBD.

Intellectual Property Statement

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

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

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

Disclaimer of Validity

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

Copyright Statement

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

Acknowledgment

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