SMTP Service Extension
for 8bit-cleanliness

Fri Oct 16 23:17:52 1992

Einar A. Stefferud
Network Management Associates, Inc.
stef@nma.com

David H. Crocker
The Branch Office
dcrocker@mordor.stanford.edu

Marshall T. Rose
Dover Beach Consulting, Inc.
mrose@dbc.mtview.ca.us

[1]. Status of this Memo

This document is an Internet Draft. 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 valid for a maximum of six months and may
be updated, replaced, or obsoleted by other documents at any
time. (The file 1id-abstracts.txt on nic.ddn.mil describes
the current status of each Internet Draft.) It is not
appropriate to use Internet Drafts as reference material or to
cite them other than as a "work in progress".

[2]. Abstract

This memo defines an extension to the SMTP service whereby an
SMTP content body containing octets outside of the US ASCII
repertoire may be relayed using SMTP.

3. Introduction

The Simple Mail Transfer Protocol (SMTP) [1] has provided a
stable, effective basis for the relay function of message
transfer agents. Although a decade old, SMTP has proven
remarkably resilient.

4. Framework for SMTP Extensions

For the purpose of service extensions to the SMTP, SMTP relays
a mail object containing an envelope and a content.

(1) The SMTP envelope is straight-forward, and is sent as a
     series of SMTP protocol units: it consists of an
     originator address (to which error reports should be
     directed); a delivery mode (e.g., deliver to recipient
     mailboxes); and, one or more recipient addresses.

(2) The SMTP content is sent in the SMTP DATA protocol unit
     and has two parts: the headers and the body. The headers
     form a collection of field/value pairs structured
     according to RFC 822 [2], whilst the body, if structured,
     is defined according to MIME [3]. The content is textual
     in nature, expressed using the US ASCII repertoire (ISO
     registered character set number 6). Although extensions
     (such as MIME) may relax this restriction for the content
     body, the content headers are always encoded using the US
     ASCII repertoire. The algorithm defined in [4] is used
     to represent header values outside the US ASCII
     repertoire, whilst still encoding them using the US ASCII
     repertoire.

Although SMTP is widely and robustly deployed, some parts of
the Internet community might wish to extend the SMTP service.
In particular, a significant portion of the Internet community
wishes to exchange messages in which the content body contains
arbitrary octet-aligned material. This memo uses the
mechanism described in [5] to define an extension to the SMTP
service whereby such contents may be exchanged:

(1) the name of the SMTP service extension is 8bit-
     cleanliness;

(2)   the EHLO keyword value associated with the extension is
      8BIT;

(3)   the syntax of the parameter (using the ABNF notation of
      [2]) associated with this EHLO keyword value is:

          8bit-param   ::= "CONVERT" / "NOCONVERT" / "MIME"

(4)   the next section specifies how support for the extension
      affects the behavior of a server and client SMTP.

5.   **The 8bit-cleanliness service extension**

When a client SMTP wishes to relay (using the MAIL command) a
content body which contains arbitrary octet-aligned material,
it first issues the EHLO command to the server SMTP. If the
server SMTP responds with code 250 to the EHLO command, and
the response includes the EHLO keyword value 8BIT, then the
server SMTP is indicating that it supports the extended MAIL
command and will accept some types of octet-aligned material.

If the parameter associated with the EHLO keyword value 8BIT
is "MIME", then the client SMTP may transmit a content body
containing octet-aligned material only if it is structured
according to the rules given in [3] (i.e., any body part
containing arbitrary octet-aligned material is tagged with a
Content-Transfer-Encoding header field).

Otherwise, if the parameter associated with the EHLO keyword
value 8BIT is "CONVERT" or "NOCONVERT", then the client SMTP
is permitted to transmit a content body containing arbitrary
octet-aligned material to the server SMTP.

The extended MAIL command is issued by a client SMTP when it
wishes to transmit a content body containing arbitrary octet-
aligned material. The syntax for this command is identical to
the MAIL command in [1], except that a BODY parameter may
appear after the address. The syntax of the extended MAIL
command (using the ABNF notation of [2]) is:

     email-cmd   ::= "MAIL FROM:<" reverse-path ">" body-param

     body-param   ::= SPACE "BODY=" ("USASCII" / "OCTET")

The value associated with the BODY parameter indicates whether
the content body which will be passed using the DATA command
contains some arbitrary octet-aligned material ("OCTET") or is
encoded entirely in US ASCII ("ASCII").

A server which supports the 8-bit cleanliness service
extension shall preserve all bits in each octet passed using
the DATA command.

Naturally, the usual SMTP data-stuffing algorithm applies so
that a content which contains the five-character sequence of

        <CR> <LF> <DOT> <CR> <LF>

or a content that begins with the three-character sequence of

        <DOT> <CR> <LF>

does not prematurely terminate the transfer of the content.
Further, it should be noted that the CR-LF pair immediately
preceeding the final dot is considered part of the content.
Finally, although the content body contains arbitrary octet-
aligned material, the length of each line (number of octets
between two CR-LF pairs), must not exceed 1000 characters, as
specified in [1].

Once a server SMTP supporting the 8bit-cleanliness service
extension accepts a content body containing octets with the
high-order 8th bit set, the server SMTP must deliver or relay
the content in such a way as to preserve all bits in each
octet.

If the path to the next hop (or a recipient's user agent) is
not known to preserve all bits in each octet, then the
behavior of the server SMTP depends on the parameter
associated with the EHLO keyword value 8BIT.

If the parameter had the value "CONVERT" or "MIME", then the
server SMTP must transform the content body (as described in
Section 5.1).

Otherwise, if the parameter had the value "NOCONVERT", then
this is treated as a permanent error, and is handled in the
usual manner for delivery failures for that server SMTP.

If a server SMTP does not support the 8-bit cleanliness
extension (either by not responding with code 250 to the EHLO
command, or by not including the EHLO keyword value 8BIT in
its response), then the client SMTP must not, under any
circumstances, attempt to transfer a content which contains
characters outside the US ASCII repertoire.

A client SMTP has two options in this case: first, it may seek
another server SMTP to act as a mail relay for the current
recipients (according to the rules of [6]), and which supports
the 8-bit cleanliness extension (this might result in the
message being re-queued by the client SMTP); or, it may
implement the gateway transformations described in Section
5.1. **If neither option is available, then this is treated as**
a permanent error, and is handled in the usual manner for
delivery failures for that server SMTP.

5.1. **Client SMTP performing gateway transformations**

If the client SMTP chooses to operate as a mail domain
gateway, then it first examines the content headers looking
for two header fields: MIME-Version and Content-Type. If both
are present, and if the content-type associated with the
latter is either Multipart or Message, then the client SMTP
recursively examines, and if necessary, transforms, the
transfer encodings of the body parts contained therein. For
example, if only one part of a multipart content-type is
encoded using characters outside of the US ASCII repertoire,
then only that part is transformed.

Otherwise, if the content-type is neither Multipart nor
Message (or lacks the MIME-Version and/or Content-Type
fields), then the client SMTP applies either the quoted-
printable or base64 transfer-encoding to the content body.
(The choice of transfer-encoding is at the discretion of the
client SMTP.) The Content-Transfer-Encoding field in the
content headers is changed to reflect the transfer-encoding in
use, as specified in [3]. (If this field is not present, then
it is added.) If a MIME-Version field is not present, then it
is added, i.e.,

     MIME-Version: 1.0

If a Content-Type field is not present, then it, along with a
Content-Description field, shall be added, specifically: i.e.,

        Content-Type: application/octet-stream
        Content-Description: untagged data converted to MIME

After transforming the content body, the client SMTP
determines if any textual components of the content headers
contain characters outside of the US ASCII repertoire. If so,
these are transformed using the rules defined in [4].
However, since no information is available as to the
repertoire in use, the charset value of "unknown" is used.
For each word encoded, the choice of applying either the "Q"
or "B" transfer-encoding lies with the client SMTP.

Finally, the client SMTP adds trace information to the content
headers. This takes the form of a single Received: field
placed above any other such fields in the content headers.
This new field should indicate the nature of the
transformation which was applied, using a "with" phrase, with
a value of "binary-to-base64" or "binary-to-quoted", e.g.,

        Received: from baiji.dbc.mtview.ca.us by dbc.mtview.ca.us
                  with binary-to-base64
               ; Tue, 01 Sep 1992 01:18:00 -0700

6. Acknowledgements

The authors acknowledge the comments of Harald T. Alvestrand,
Ned Freed, John C. Kleinsen, Keith Moore, Neil W. Rickert,
and, John Wagner.

7. References

[1] J.B. Postel. Simple Mail Transfer Protocol. Request for
     Comments 821, (August, 1982).

[2] D.H. Crocker. Standard for the Format of ARPA Internet
     Text Messages. Request for Comments 822, (August, 1982).

[3] N.S. Borenstein, N. Freed. Multipurpose Internet Mail
     Extensions. Request for Comments 1341, (June, 1992).

[4] K. Moore. Representation of Non-ASCII Text in Internet
     Message Headers. Request for Comments 1342, (June,
     1992).

[5] M.T. Rose, E.A. Stefferud, D.H. Crocker. SMTP Service
     Extensions. Internet-Draft, (September, 1992).

[6] C. Partridge. Mail Routing and the Domain System.
     Request for Comments 974, (January, 1986).