                     The Content-MD5 Header Field

Status of this Memo

Abstract

   This memo specifies an optional header field, Content-MD5, for use
   with MIME-conformant messages.

Table of Contents

1. Introduction

   Despite all of the mechanisms provided by MIME [1] which attempt to
   protect data from being damaged in the course of email transport, it
   is still desirable to have a mechanism for verifying that the data,
   once decoded, are intact. For this reason, this memo defines the use
   of an optional header field, Content-MD5, which may be used as a
   message integrity check (MIC), to verify that the decoded data are
   the same data that were initially sent.

   MD5 is an algorithm for computing a 128 bit "digest" of arbitrary-
   length data, with a high degree of confidence that any alterations in
   the data will be reflected in alterations in the digest. The MD5
   algorithm itself is defined in [2]. This memo specifies how the
   algorithm may be used as an integrity check for MIME mail.

2. Generation of the Content-MD5 Field

   The Content-MD5 field is generated by only an originating user agent.
   Message relays and gateways are expressly forbidden from generating a
   Content-MD5 field.

   Use of the Content-MD5 field is completely optional, but its use is
   recommended whenever data integrity is desired, but Privacy-Enhanced
   Mail services [3] are not available. (Consult Section 4 for further
   details.) The Content-MD5 field may only be added to MIME entities of
   a 'leaf' nature, i.e., the Content-MD5 field may be used with any
   content type other than multipart or message/rfc822.

   To generate the value of the Content-MD5 field, the MD5 algorithm is
   computed on the canonical form of the data. In particular, this
   means that the sender applies the MD5 algorithm on the raw data,
   before applying any content-transfer-encoding, and that the receiver
   also applies the MD5 algorithm on the raw data, after undoing any
   content-transfer-encoding. For textual data, the MD5 algorithm must
   be computed on data in which the canonical form for newlines applies,
   that is, in which each newline is represented by a CR-LF pair.

   The output of the MD5 algorithm is a 128 bit digest. When viewed in
   network byte order (big-endian order), this yields a sequence of 16
   octets of binary data. These 16 octets are then encoded according to
   the base64 algorithm in order to obtain the value that is placed in
   the Content-MD5 field. Thus, if the application of the MD5 algorithm
   over the raw data of a MIME entity results in a digest having the
   (unlikely) value of "Check Integrity!", then that MIME entity's
   header could contain the field

           Content-MD5:   Q2hlY2sgSW50ZWdyaXR5IQ==

   Finally, as discussed in Appendix B of [1], textual data is regularly
   altered in the normal delivery of mail. Because the addition or
   deletion of trailing white space will result in a different digest,
   either the quoted-printable or base64 algorithm should be employed as
   a content-transfer-encoding when the Content-MD5 field is used.

3. Processing the Content-MD5 field

   If the Content-MD5 field is present, a recipient user agent may
   choose to use it to verify that the contents of a MIME entity have
   not been modified during transport. Message relays and gateways are
   expressly forbidden to alter its processing based on the presence of
   the Content-MD5 field. However, a message gateway is allowed to
   remove the Content-MD5 field if the corresponding MIME entity is
   translated into a different content-type.

4. Security Considerations

   This document specifies a data integrity service that protects data
   from accidental modification while in transit from the sender to the
   recipient. A secure data integrity service, such as that provided by
   Privacy Enhanced Mail [3], is conjectured to protect data from all
   modifications.

5. Acknowledgements

   This memo is based almost entirely on text originally written by
   Nathaniel Borenstein of Bellcore. In addition, several improvements
   were suggested by Keith Moore of the University of Tennessee,
   Knoxville.

6. References

   [1] Borenstein, N., and N. Freed, "MIME (Multipurpose Internet Mail
       Extensions) Part One: Mechanisms for Specifying and Describing
       the Format of Internet Message Bodies", RFC 1521, Bellcore,
       Innosoft, September 1993.

   [2] Rivest, R., "The MD5 Message-Digest Algorithm", RFC 1321, MIT
       Laboratory for Computer Science and RSA Data Security, Inc.,
       April 1992.

   [3] Linn, J., "Privacy Enhancement for Internet Electronic Mail, Part
       I: Message Encryption and Authentication Procedures", RFC 1421,
       IAB IRTF PSRG, IETF PEM WG, February 1993.

7. Author's Address

   Marshall T. Rose
   Dover Beach Consulting, Inc.
   420 Whisman Court
   Mountain View, CA 94043-2112

   Phone: (415) 968-1052
   EMail: mrose@dbc.mtview.ca.us