Content-ID and Message-ID Uniform Resource Locators

This draft document is being circulated for comment. Please send your
comments to the authors or to the uri mail list <uri@bunyip.com>. If
consensus is reached, this Access Type may be registered with IANA and
this document may be submitted to the RFC editor as an Informational
protocol specification.

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
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.
They may be updated, replaced, or obsoleted by other documents at any
time. It is not appropriate to use Internet Drafts as reference
material or to cite them other than as a "working draft" or "work in
progress".

Please check the abstract listing in each Internet Draft directory for
the current status of this or any other Internet Draft.

Abstract

The Uniform Resource Locator (URL) schemes, "cid:" and "mid:" allow the
content of Text/HTML or other MIME media types to contain references to
other body parts in the same or a different message.

0. Changes from 01

Corrected cid in example.

Changed separator in cidmid to "/"

Revised relationship between mid and cid urls and Message- and Content-
IDs.

Added rule for choosing among several body parts with the same Content-
ID.

1. Introduction

There is a desire to refer to email or NetNews messages and the body
parts of such messages using Uniform Resource Locators (URLs). The MID
and CID schemes described below permit such references.

The use of [MIME] within email to convey Web pages and their associated
images requires a URL scheme to permit the HTML to refer to the images
included in the message. A Content-ID Uniform Resource Locator serves
that purpose. Similarly NetNews readers use Message-IDs to link related
messages together. The Message-ID URL provides a scheme to to refer to
such messages as a "resource".

The MID and CID URLs represent identifiers for messages and their body
parts. In the case of a MID the message must exist inside of the user's
mail storage. A CID URL refers to a body part within the same message
as a referring body body. A CID may occur as part of a MID in which
case the CID refers to a body part in the message identified by the MID
URL portion.

A note on terminology. The terms "body part" and "MIME entity" are used
interchangeably. They refer to the headers and body of a MIME message,
either the message itself or one of the body parts contained in a
Multipart message.

2. The MID and CID URL Schemes

RFC1738 [URL] reserves the "mid" and "cid" schemes for Message-ID and
Content-ID respectively. This memorandum defines the syntax for those
URLs. Because they use the same syntactic elements they are presented
together.

The URLs takes the form

      cidurl   = "cid" ":" addr-spec

      midurl   = msgmid / cidmid

      msgmid   = "mid" ":" addr-spec

      cidmid   = msgmid "/" addr-spec

where "addr-spec" is defined in [822]. The midurl and cidurl
"schemeparts" must consist of an "xchar" sequence [URL]. Addr-spec,
however, admits a wider range of characters. Consequently some charac-
ters in addr-spec must be represented within a midurl or cidurl using
the escape mechanism in [URL].

      NOTE: Various separators have been suggested, based on an analogous

use in other URLs, for the cidmid rule: slash, "/"; question mark,
"?"; and number sign, "#". The slash suggests a hierarchical rela-
tionship between the cidurl and midurl, analogous to file system
directories; the questions mark, that the cidurl is a search argu-
ment; and the number sign, that the cidurl is a label, analogous an
anchor identifier. The slash was chosen to avoid encoding the num-
ber sign and avoid suggesting that a search was required.

The analogy to a hierarchy should not be taken strictly. A midurl
reference may be a message with a MIME entity and corresponding
Content-ID that contains another message (Content-Type: Mes-
sage/822) with its own Message-ID and that enclosed message may
have entities labeled with Content-ID. The midurl syntax, to avoid
any need for recursion, intentionally does not support that situa-
tion. To be referenced by a midurl, an enclosed message but be
extracted and be directly represented in the users message system.

A msgmid refers to the entire message and the cidmid refers to a single
body part within the referenced message. A cidurl refers to another
body part within the message that contains the cidurl.

A message may contain, usually in a Multipart/Alternate, several bodies
with the same Content-ID. A cidurl reference may thus be ambiguous; the
Multipart/Alternate [MIME] selection rules shall apply to disambiguate
the referenced body part.

A msgmid (cidurl) can be converted to its corresponding Message-ID (Con-
tent-ID) by removing the "mid:" ("cid:) prefix, converting escaped char-
acters to their ASCII equivalent, and enclosing the remaining part with
an angle bracket pair, "<" and ">". Ignoring the escape mechanism,
"mid:_addr-spec_" has the message-id "<_addr-spec_>". Similarly, a cid-
mid can be converted to a message-id, content-id pair.

Cidurl and midurls (message-ids and content-ids) are globally unique
[MIME, p.19]. A common technique for generating a globally unique
cidurl and midurl uses a time and date stamp with the local host's
domain name, e.g., 950124.162336@XIson.com.

3. Security

Security issues are not addressed in this memorandum.

4. References

[822]
    Crocker, D., Standard for the Format of ARPA Internet Text Mes-
    sages, August 1982, University of Delaware, RFC 822, STD 11.

[MIME]
     N. Borenstein, N. Freed, "MIME (Multipurpose Internet Mail Exten-
     sions) Part One: Mechanisms for Specifying and Describing the For-
     mat of Internet Message Bodies", 09/23/1993, RFC 1521.

[URL]
     Berners-Lee, T., Masinter, L., and McCahill, M., Uniform Resource
     Locators (URL), December 1994, RFC 1738.

5. Acknowledgements

This work reflects the ideas freely provided to the author by Harald T.
Alvestrand, UNINETT, including Tim Berners-Lee, W3O, who pointed me at
the idea of using a URL "scheme" in the SGML encapsulation proposal,
Daniel W. Connolly, HAL, and Roy T. Fielding, UCI.

6. Author's Address

Edward Levinson
XIson, Inc.
47 Clive Street
Metuchen, NJ 08840-1060
USA
+1 908 549 3716
<XIson@cnj.digex.net>