

Content-ID and Message-ID Uniform Resource Locators

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (1998). All Rights Reserved.

Abstract

The Uniform Resource Locator (URL) schemes, "cid:" and "mid:" allow references to messages and the body parts of messages. For example, within a single multipart message, one HTML body part might include embedded references to other parts of the same message.

Changes from (RFC 2111)

Clarified the example on page 3 on of converting cid URLs to Content-IDs. The example now uses a cid URL instead of an mid.

Corrected the example messages to have the correct Content-ID form; they now use the angle brackets. Added a Message-ID header to the second example.

1. Introduction

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 or other data included in the message. The Content-ID Uniform Resource Locator, "cid:", serves that purpose.

Similarly Net News readers use Message-IDs to link related messages together. The Message-ID URL provides a scheme, "mid:", to refer to such messages as a "resource".

The "mid" (Message-ID) and "cid" (Content-ID) URL schemes provide identifiers for messages and their body parts. The "mid" scheme uses (a part of) the message-id of an email message to refer to a specific message. The "cid" scheme refers to a specific body part of a message; its use is generally limited to references to other body parts in the same message as the referring body part. The "mid" scheme may also refer to a specific body part within a designated message, by including the content-ID's address.

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

RFC 1738 [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 take the form

content-id = url-addr-spec

message-id = url-addr-spec

url-addr-spec = addr-spec ; URL encoding of RFC 822 addr-spec

cid-url = "cid" ":" content-id

mid-url = "mid" ":" message-id ["/" content-id]

Notes: In Internet mail messages, the addr-spec in a Content-ID [MIME] or Message-ID [822] header is enclosed in angle brackets (<>). Since addr-spec in a Message-ID or Content-ID might contain characters not allowed within a URL; any such character (including "/", which is reserved within the "mid" scheme) must be hex-encoded using the %hh escape mechanism in [URL].

A "mid" URL with only a "message-id" refers to an entire message. With the appended "content-id", it refers to a body part within a message, as does a "cid" URL. The Content-ID of a MIME body part is required to be globally unique. However, in many systems that store messages, body parts are not indexed independently their context (message). The "mid" URL long form was designed to supply the context needed to support interoperability with such systems.

A implementation conforming to this specification is required to support the "mid" URL long form (message-id/content-id). Conforming implementations can choose to, but are not required to, take advantage of the content-id's uniqueness and interpret a "cid" URL to refer to any body part within the message store.

In limited circumstances (e.g., within multipart/alternate), a single message may contain several body parts that have the same Content-ID. That occurs, for example, when identical data can be accessed through different methods. In those cases, conforming implementations are required to use the rules of the containing MIME entity (e.g., multipart/alternate) to select the body part to which the Content-ID refers.

A "cid" URL is converted to the corresponding Content-ID message header [MIME] by removing the "cid:" prefix, converting the % encoded character to their equivalent US-ASCII characters, and enclosing the remaining parts with an angle bracket pair, "<" and ">". For example, "cid:foo4%25foo1@bar.net" corresponds to

Content-ID: <foo4%25foo1@bar.net>

Reversing the process and converting URL special characters to their % encodings produces the original cid.

A "mid" URL is converted to a Message-ID or Message-ID/Content-ID pair in a similar fashion.

Both message-id and content-id are required to be globally unique. That is, no two different messages will ever have the same Message-ID addr-spec; no different body parts will ever have the same Content-ID addr-spec. A common technique used by many message systems is to use a time and date stamp along with the local host's domain name, e.g., 950124.162336@Xison.com.

Some Examples

The following message contains an HTML body part that refers to an image contained in another body part. Both body parts are contained in a Multipart/Related MIME entity. The HTML IMG tag contains a cidurl which points to the image.

```
From: foo1@bar.net
To: foo2@bar.net
Subject: A simple example
Mime-Version: 1.0
Content-Type: multipart/related; boundary="boundary-example-1";
              type=Text/HTML
```

--boundary-example 1

Content-Type: Text/HTML; charset=US-ASCII

to the other body part, for example through a statement such as:

--boundary-example-1

Content-ID: <foo4*foo1@bar.net>

Content-Type: IMAGE/GIF

Content-Transfer-Encoding: BASE64

R0lGODlhGAGgAPEAAP/////ZRACgoAAAACH+PUNvcHlyaWdodCAoQykgMTk5
 NSBJRVRGLiBVbmF1dGhvcml6ZWQgZHVwbGljYXRpb24gcHJvaGliaXRlZC4A
 etc...

--boundary-example-1--

The following message points to another message (hopefully still in the recipient's message store).

From: bar@none.com

To: phooey@all.com

Subject: Here's how to do it

Message-ID: <970701.32784@VIers.none.com>

Content-type: text/html; charset=usascii

 previous message, shows how the approach you propose can be
 used to accomplish ...

3. Security Considerations

The URLs defined here provide an addressing or referencing mechanism. The values of these URLs disclose no more about the originators environment than the corresponding Message-ID and Content-ID values. Where concern exists about such disclosures the originator of a message using mid and cid URLs must take precautions to insure that confidential information is not disclosed. Those precautions should already be in place to handle existing mail use of the Message-ID and Content-ID.

4. References

- [822] Crocker, D., "Standard for the Format of ARPA Internet Text Messages", August 1982, STD 11, RFC 822, August 1982.
- [MIME] Borenstein, N., and N. Freed, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [URL] Berners-Lee, T., Masinter, L., and M. McCahill, "Uniform Resource Locators (URL)", RFC 1738, December 1994.
- [MULREL] Levinson, E., "The MIME Multipart/Related Content-type", RFC 2387, August 1998.

5. Acknowledgments

The original concept of "mid" and "cid" URLs were part of the Tim Berners-Lee's original vision of the World Wide Web. The ideas and design have benefited greatly by discussions with Harald Alvestrand, Dan Connolly, Roy Fielding, Larry Masinter, Jacob Palme, and others in the MHTML working group.

6. Author's Address

Edward Levinson
47 Clive Street
Metuchen, NJ 08840-1060
USA

Phone: +1 908 549 3716
EMail: XIson@cnj.digex.net

7. Full Copyright Statement

Copyright (C) The Internet Society (1998). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS 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.

