

The Archived-At Message Header Field

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.

Abstract

This memo defines a new email header field, Archived-At:, to provide a direct link to the archived form of an individual email message.

Table of Contents

1. Introduction	2
2. Header Field Definition	2
2.1. Syntax	2
2.2. Multiple Archived-At Header Fields	3
2.3. Interaction with Message Fragmentation and Reassembly	3
2.4. Syntax Extension for Internationalized Message Headers	3
2.5. The X-Archived-At Header Field	4
3. Implementation and Usage Considerations	4
3.1. Formats of Archived Message	4
3.2. Implementation Considerations	4
3.3. Usage Considerations	5
4. Security Considerations	6
5. IANA Considerations	7
5.1. Registration of the Archive-At Header Field	7
5.2. Registration of the X-Archived-At Header Field	7
6. Acknowledgments	8
7. References	8
7.1. Normative References	8
7.2. Informative References	8

1. Introduction

[RFC2369] defines a number of header fields that can be added to Internet messages such as those sent by email distribution lists or in netnews [RFC1036]. One of them is the List-Archive header field that describes how to access archives for the list. This allows access to the archives as a whole, but not an individual message.

There is often a need or desire to refer to the archived form of a single message. For more detailed usage scenarios, please see Section 3.3. This memo defines a new header, Archived-At, to refer to a single message at an archived location. This provides quick access to the location of a mailing list message in the list archive. It can also be used independently of mailing lists, for example in connection with legal requirements to archive certain messages.

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in [RFC2119].

2. Header Field Definition

2.1. Syntax

For the Archived-At header field, the field name is "Archived-At". The field body consist of a URI [STD66] enclosed in angle brackets ('<', '>'). The URI MAY contain folding whitespace (FWS, [RFC2822]), which is ignored. Mail Transfer Agents (MTAs) MUST NOT insert whitespace within the angle brackets, but client applications SHOULD ignore any whitespace, which might have been inserted by poorly behaved MTAs. The URI points to an archived version of the message. See Section 3.1 for more details.

This header field is subject to the encoding and character restrictions for mail headers as described in [RFC2822].

More formally, the header field is defined as follows in Augmented BNF (ABNF) according to [RFC4234]:

```
archived-at = "Archived-At:" [FWS] "<" folded-URI ">" CRLF
folded-URI  = <URI, but free insertion of FWS permitted>
```

where URI is defined in [STD66], and CRLF and FWS are defined in [RFC2822].

To convert a folded-URI to a URI, first apply standard [RFC2822] unfolding rules (replacing FWS with a single SP), and then delete any remaining un-encoded SP characters.

This syntax is kept simple in that only one URI per header field is allowed. In this respect, the syntax is different from [RFC2369]. Also, comments are not allowed.

2.2. Multiple Archived-At Header Fields

Each Archived-At header field only contains a single URI. If it is desired to list multiple URIs where an archived copy of the message can be found, a separate Archived-At field per URI is required. Multiple Archived-At header fields with the same URI SHOULD be avoided. An Archived-At header field SHOULD only be created if the message is actually being made available at the URI given in the header field.

If a message is forwarded from a list to a sublist and both lists support adding the Archived-At header field, then the sublist SHOULD add a new Archived-At header field without removing the already existing one(s), unless the header field is exactly the same as an already existing one, in which case the new header field SHOULD NOT be added.

2.3. Interaction with Message Fragmentation and Reassembly

[RFC2046] allows for the fragmentation and reassembly of messages. Archived-At header fields are to be treated in the same way as Comments header fields, i.e., copied to the first fragment message header on fragmentation and back from there to the header of the reassembled message.

This treatment has been chosen for compatibility with existing infrastructure. It means that Archived-At header fields in the first fragment message MAY refer to an archived version of the whole, unfragmented message. To avoid confusion, Archived-At headers SHOULD NOT be added to fragment messages.

2.4. Syntax Extension for Internationalized Message Headers

There are some efforts to allow non-ASCII text directly in message header field bodies. In such contexts, the URI non-terminal in the syntax defined in Section 2.1 is to be replaced by an Internationalized Resource Identifier (IRI) as defined in [RFC3987]. The specifics of the actual octet encoding of the IRI will follow the rules for general direct encoding of non-ASCII text. For conversion between IRIs and URIs, the procedures defined in [RFC3987] are to be applied.

2.5. The X-Archived-At Header Field

For backwards compatibility, this document also describes the X-Archived-At header field, a precursor of the Archived-At header field. The X-Archived-At header field MAY also be parsed, but SHOULD not be generated.

The following is the syntax of the X-Archived-At header field in ABNF according to [RFC4234] (which also defines SP):

```
obs-archived-at = "X-Archived-At:" SP URI CRLF
```

The X-Archived-At header field does not allow whitespace inside URI.

3. Implementation and Usage Considerations

3.1. Formats of Archived Message

There is no restriction on the format used to serve the archived message from the URI in an Archived-At header field. It is expected that in many cases, the archived message will be served as (X)HTML, as plain text, or in its original form as message/rfc822 [RFC2046]. Some forms of URIs may imply the format in which the archived message is served, although this should not be relied upon.

If the protocol used to retrieve the message allows for content negotiation, then it is also possible to serve the archived message in several different formats. As an example, an HTTP URI in an Archived-At header may make it possible to serve the archived message both as text/html for human consumption in a browser and as message/rfc822 for use by a mail user agent (MUA) without loss of information.

3.2. Implementation Considerations

Mailing list expanders and email archives are often separate pieces of software. It may therefore be difficult to create an Archived-At header field in the mailing list expander software.

One way to address this difficulty is to have the mailing list expander software generate an unambiguous URI, e.g., a URI based on the message identifier of the incoming email, and to set up the archiving system so that it redirects requests for such URIs to the actual messages. If the email does not contain a message identifier, a unique identifier can be generated.

Such a system has been implemented and is in productive use at W3C. As an example, the URI "http://www.w3.org/mid/0I5U00G08DFGCR@mailsj-v1.corp.adobe.com", containing the significant part of the message identifier "<0I5U00G08DFGCR@mailsj-v1.corp.adobe.com>", is redirected to the URI of this message in the W3C mailing-list archive at <http://lists.w3.org/Archives/Public/uri/2004Oct/0017.html>.

Source code for this implementation is available at <http://dev.w3.org/cvsweb/search/>, in particular <http://dev.w3.org/cvsweb/search/cgi/mid.pl> and <http://dev.w3.org/cvsweb/search/bin/msgid-db.pl>. These locations may be subject to change.

When using the message identifier to create an address for the archived mail, care has to be taken to escape characters in the message identifier that are not allowed in the URI, or to remove them, as done above for the "<" and ">" delimiters.

Implementations such as that described above can introduce a security issue. Somebody might deliberately reuse a message identifier to break the link to a message. This can be addressed by checking incoming message identifiers against those of the messages already in the archive and discarding incoming duplicates, by checking the content of incoming duplicates and discarding them if they are significantly different from the first message, by offering multiple choices in the response to the URI, or by using some authentication mechanism on incoming messages.

3.3. Usage Considerations

It may at first seem strange to have a pointer to an archived form of a message in a header field of that same message. After all, if one has the message, why would one need a pointer to it? It turns out that such pointers can be extremely useful. This section describes some of the scenarios for their use.

A user may want to refer to messages in a non-message context, such as on a Web page, in an instant message, or in a phone conversation. In such a case, the user can extract the URI from the Archived-At header field, avoiding the search for the correct message in the archive.

A user may want to refer to other messages in a message context. Referring to a single message is often done by replying to that message. However, when referring to more than one message, providing pointers to archived messages is a widespread practice. The Archived-At header field makes it easier to provide these pointers.

A user may want to find messages related to a message at hand. The user may not have received the related messages, and therefore needs to use an archive. The user may also prefer finding related messages in the archive rather than in her MUA, because messages in archives may be linked in ways not provided by the MUA. The Archived-At header field provides a link to the starting point in the archive from which to find related messages.

Please note that in the above usage scenarios, it is mostly the human reader, rather than the email client software, that makes use of the URI in the Archived-At header. However, this does not rule out the use of the URI in the Archived-At header by the email client or other software if such use is found helpful.

4. Security Considerations

There are many potential security issues when activating and dereferencing a URI. For more details, including some countermeasures, please see [STD66]. In the context of this proposal, the following are particularly relevant: An intruder may get access to the message transmission and be able to insert a URI pointing to some malicious content. This can be addressed by using a secured way of message transmission. Also, somebody may be able to construct a message that is harmless when received directly, but that produces problems when accessed via the URI. One reason for this may be the format used in the archive, where some content was not adequately escaped. This can be addressed by using adequate escaping.

The Archived-At header field points to some archived form of the message itself. This in turn may contain the Archived-At field. This creates a potential for a denial-of-service attack on the server pointed to by the URI in the Archived-At header field. The conditions are that the archived form of the message is downloaded automatically, and that further URIs in that message are followed and downloaded recursively without checking for already downloaded resources. However, this kind of scenario can easily be avoided by implementations. First, the URI in the Archived-At header field should not be dereferenced automatically. Second, appropriate measures for loop detection should be used.

In Section 3.2, an attack is described that may break a URI to a message by introducing a new message with the same message identifier. Possible countermeasures are also discussed.

5. IANA Considerations

5.1. Registration of the Archive-At Header Field

IANA has registered the Archived-At header field in the Message Header Fields Registry ([RFC3864]) as follows:

Header field name:
 Archived-At

Applicable protocol:
 mail (RFC 2822) and netnews (RFC 1036)

Status:
 standard

Author/Change controller:
 IETF

Specification document(s):
 RFC 5064

Related information:
 none

5.2. Registration of the X-Archived-At Header Field

This section is non-normative (specifically, an implementation that ignores this section remains compliant with this specification).

IANA has registered the X-Archived-At header field in the Message Header Fields Registry ([RFC3864]) as follows:

Header field name:
 X-Archived-At

Applicable protocol:
 mail (RFC 2822) and netnews (RFC 1036)

Status:
 deprecated

Author/Change controller:
 IETF

Specification document(s):
 RFC 5064

Related information:
none

6. Acknowledgments

The members of the W3C system team, in particular Gerald Oskoboiny, Olivier Thereaux, Jose Kahan, and Eric Prud'hommeaux, created the mid-based email archive lookup system and the experimental form of the Archived-At header. Pete Resnik provided the motivation for writing this memo. Discussion on the ietf-822@imc.org mailing list, in particular contributions by Frank Ellermann, Arnt Gulbrandsen, Graham Klyne, Bruce Lilly, Charles Lindsey, and Keith Moore, led to further improvements of the proposal. Chris Newman, Chris Lonvick, Stephane Borzmeyer, Vijay K. Gurbani, and S. Moonesamy provided additional valuable comments.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2822] Resnick, P., Ed., "Internet Message Format", RFC 2822, April 2001.
- [RFC3864] Klyne, G., Nottingham, M., and J. Mogul, "Registration Procedures for Message Header Fields", BCP 90, RFC 3864, September 2004.
- [RFC3987] Duerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)", RFC 3987, January 2005.
- [RFC4234] Crocker, D., Ed., and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 4234, October 2005.
- [STD66] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

7.2. Informative References

- [RFC1036] Horton, M. and R. Adams, "Standard for interchange of USENET messages", RFC 1036, December 1987.
- [RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, November 1996.

[RFC2369] Neufeld, G. and J. Baer, "The Use of URLs as Meta-Syntax for Core Mail List Commands and their Transport through Message Header Fields", RFC 2369, July 1998.

Author's Address

Martin Duerst (Note: Please write "Duerst" with u-umlaut wherever possible, for example as "Dürst" in XML and HTML.)
Aoyama Gakuin University
5-10-1 Fuchinobe
Sagamihara, Kanagawa 229-8558
Japan

Phone: +81 42 759 6329

Fax: +81 42 759 6495

EMail: duerst@it.aoyama.ac.jp

URI: <http://www.sw.it.aoyama.ac.jp/D%C3%BCrst/>

Full Copyright Statement

Copyright (C) The IETF Trust (2007).

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.

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, THE IETF TRUST 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.

Intellectual Property

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.