

Mbox

Mbox is a generic term for a family of related file formats used for holding collections of email messages, first implemented for Fifth Edition Unix.

All messages in an mbox mailbox are concatenated and stored as plain text in a single file. Each message starts with the four characters "From" followed by a space (the so named "From_ line") and the sender's email address. RFC 4155 defines that a UTC timestamp follows after another separating space character.

Unlike the Internet protocols used for the exchange of email, the format used for the storage of email has never been formally defined through the RFC standardization mechanism and has been entirely left to the developer of an email client. However, the POSIX standard defined a loose frame in conjunction with the mailx program. In 2005 finally, the application/mbox media type was standardized as RFC 4155, and hints that mbox stores mailbox messages in their original Internet Message (RFC 2822) format, except for the used newline character, seven-bit clean data storage, and the requirement that each newly added message is terminated with a completely empty line within the mbox database.

A format similar to mbox is the MH Message Handling System. Other systems, such as Microsoft Exchange Server and the Cyrus IMAP server store mailboxes in centralised databases managed by the mail system and not directly accessible by individual users. The maildir mailbox format is often cited as an alternative to the **mbox** format for network email storage systems.

Contents

Family

Modified mbox

File locking

As a patch format

See also

References

Further reading

Family

The mbox format uses a single blank line followed by the string 'From ' (with a space) to delimit messages; this can create ambiguities if a message contains the same sequence in the message text.

Over the decades that followed, four popular but incompatible variants arose: *mboxo*, *mboxrd*, *mboxcl*, and *mboxcl2*. The naming scheme was developed by Daniel J. Bernstein, Rahul Dhesi, and others in 1996. Each originated from a different version of Unix. *mboxcl* and *mboxcl2* originated from the file format used by Unix System V Release 4 mail tools. *mboxrd* was invented by Rahul Dhesi et al. as a rationalisation of *mboxo* and subsequently adopted by some Unix mail tools including gmail.

All these variants have the problem that the content of the message is modified in order to remove the ambiguities, as shown below, so that applications have to know which quoting rule has been used in order to perform the correct reversion, which turned out to be impractical. Using MIME and choosing a content-transfer-encoding that quotes "From_" lines in a standard-compliant fashion ensures that message content doesn't need to be changed, but only their MIME representation. Therefore checksums remain constant, a necessary precondition for supporting S/MIME and Pretty Good Privacy. Applications which newly create messages and store them in mbox database files will likely use this approach to detach message content from database storage format.

mboxo and *mboxrd* locate the message start by scanning for *From* lines that are found before the email message headers. If a "From " string occurs at the beginning of a line in either the header or the body of a message (a mail standard violation for the former, but not for the latter), the email message must be modified before the message is stored in an **mbox** mailbox file or the line will be taken as a message boundary. To avoid misinterpreting a "From " string at the beginning of the line in the email body as the beginning of a new email, some systems "**From-munge**"^[1] the message, typically by prepending a greater-than sign:

```
>From my point of view...
```

In the *mboxo* format, such lines have irreversible ambiguity.^[2] In the *mboxo* format, this can lead to corruption of the message. If a line already contained >From at the beginning (such as in a quotation), it is unchanged when written. When subsequently read by the mail software, the leading > is erroneously removed. The *mboxrd* format solves this by converting From to >From and converting >From to >>From, etc. The transformation is then always reversible.^[3]

Example:

```
From MAILER-DAEMON Fri Jul 8 12:08:34 2011
From: Author <author@example.com>
To: Recipient <recipient@example.com>
Subject: Sample message 1

This is the body.
>From (should be escaped).
There are 3 lines.

From MAILER-DAEMON Fri Jul 8 12:08:34 2011
From: Author <author@example.com>
To: Recipient <recipient@example.com>
Subject: Sample message 2

This is the second body.
```

The *mboxcl* and *mboxcl2* formats use a **Content-Length:** header to determine the messages' lengths and thereby the next *real From line*. *mboxcl* still quotes *From lines* in the messages themselves as *mboxrd* does, while *mboxcl2* doesn't.

Modified mbox

Some email clients use a modification of the mbox format for their mail folders.

- Eudora used an *mboxo* variation where a sender's email address is replaced by the constant string "???@???". Most mbox clients store incoming messages as received. Eudora separates out

attachments embedded in the message, storing the attachments as separate individual files in one folder.^[4]

- The Mozilla family of email clients (Mozilla, Netscape, Thunderbird, et al.) use an *mboxrd* variation with more complex *From line* quoting rules.^[5]

File locking

Various mutually incompatible mechanisms have been used by different mbox formats to enable message file locking, including `fcntl()` and `lockf()`. This does not work well with network mounted file systems, such as the Network File System (NFS), which is why traditionally Unix used additional "dot lock" files, which could be created atomically even over NFS.

Because more than one message is stored in a single file, some form of file locking is needed to avoid the corruption that can result from two or more processes modifying the mailbox simultaneously. This could happen if a network email delivery program delivers a new message at the same time as a mail reader is deleting an existing message.

Mbox files should be locked also while they are being read. Otherwise the reader may see corrupted message contents if another process is modifying the mbox at the same time, even though no actual file corruption occurs.

As a patch format

In open source development, it is common to send patches in the diff format to a mailing list for discussion. The diff format allows for irrelevant "headers", such as mbox data, to be added.^{[6][7]} Version control systems like git have support for generating mbox-formatted patches and for sending them to the list as emails in a thread.^{[8][9]}

See also

- Maildir
- MIX (email)
- MH Message Handling System

References

1. RFC 3676 "The Text/Plain Format and DelSp Parameters". Section 4.4: Space-Stuffing. 2004.
2. "Configuring Netscape Mail On Unix: Why the Content-Length Format is Bad" (<https://www.jwz.org/doc/content-length.html>) by Jamie Zawinski 1997
3. de Boyne Pollard, Jonathan (2004). "'mbox' is a family of several mutually incompatible mailbox formats" (<http://jdeb.eu/FGA/mail-mbox-formats.html>). *Frequently Given Answers*.
4. "Eudora 6.2.4 Mac User Guide" (https://web.archive.org/web/20140712122540/http://www.eudora.com/download/eudora/mac/6.2.4/Eudora_6.2.4_Mac_User_Guide.pdf) (PDF). p. 113. Archived from the original (http://www.eudora.com/download/eudora/mac/6.2.4/Eudora_6.2.4_Mac_User_Guide.pdf) (PDF) on 2014-07-12. Retrieved 2015-10-29.
5. "Importing and exporting your mail" (http://kb.mozillazine.org/Importing_and_exporting_your_mail#Mbox_files)

6. "Submitting patches: the essential guide to getting your code into the kernel — The Linux Kernel documentation" (<https://www.kernel.org/doc/html/v4.17/process/submitting-patches.html>). www.kernel.org.
7. Randal, Allison; Sugalski, Dan; Tötsch, Leopold (2003). "Patch submission". *Perl 6 Essentials* (<https://archive.org/details/perl6essentials0000rand>). O'Reilly Media, Inc. p. 14 (<https://archive.org/details/perl6essentials0000rand/page/14>). ISBN 978-0-596-00499-6.
8. "Git - git-format-patch Documentation" (<https://git-scm.com/docs/git-format-patch>). git-scm.com.
9. "Git - git-send-email Documentation" (<https://git-scm.com/docs/git-send-email>). git-scm.com.

Further reading

- Jonathan de Boyne Pollard (2004). "'mbox' is a family of several mutually incompatible mailbox formats" (<https://jdeb.eu/FGA/mail-mbox-formats.html>). *Frequently Given Answers*.
 - gmail mbox manual page (<http://gmail.org/gmail-manual-html/man5/mbox.html>)
 - Internet Mail Consortium (<https://web.archive.org/web/20120529014601/http://www.imc.org/>) – Standards body
 - Know What is MBOX File (<http://www.whatis.ibidinfo.com/file-extension/mbox-file.html>)
 - RFC 4155 – The application/mbox Media Type
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Mbox&oldid=948878742>"

This page was last edited on 3 April 2020, at 15:09 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.