What is UNIX mbox (mailbox) format?

Contents

Introduction

This document describes Unix mailbox (mbox) format and how it relates to use on the Cisco Email Security Appliance (ESA).

What is UNIX mbox (mailbox) format?

UNIX mbox format is used by AsyncOS when messages are archived and logged in the message filter log() action. "Archive Message" is an additional configuration option for Ironport Anit-spam (IPAS), Anti-virus (Sophos and McAfee), Advanced Malware Protection (AMP), and Graymail on the ESA.

Mbox format is an ASCII-formatted (that is, not binary) file format that can contain zero or more mail messages. Messages are concatenated in the mbox file and can be pried apart based on specific strings in the file. This format is identical with the message as they are transferred between RFC 2821 compliant mail gateways.

Each message in mbox format begins with a line that starts with the string "From" (ASCII characters F, r, o, m, and space). "From" lines are followed by several more fields: envelope-sender, date, and (optionally) more data.

The first field after the "From" string is the envelope-sender of the message. Dependent upon which application creates the mbox file, the envelope-sender might be present as a real mailbox or it might be another character or string. Most commonly, you will find that a "-" (single character dash) replaces the envelope-sender if the actual envelope-sender is not available or not known. The date field inserted by the ESA is in standard UNIX asctime() format and is always 24 characters in length. In some mbox files written by non-AsyncOS implementations, further information follows the date stamp. These three fields are separated by a single space.

Here is an example of an mbox file with a single message in it:

```
From Adam@Outside.COM Sun Oct 17 12:03:20 2004
Received: from mail.outside.com (192.35.195.200)
by smtp.alpha.com with ESMTP; 17 Oct 2004 12:03:20 -0700
X-IronPort-AV: i="3.85,147,1094454000";
v="EICAR-AV-Test'0'v";
d="scan'208"; a="86:adNrHT37924848"
X-IronPort-RCPT-TO: alan@mail.example.com
From: Adam@Outside.COM
To: Alan Alpha <Alan@mail.example.COM>
Subject: Exercise 7a Anti-Virus Scanning
Reply-To: Adam Alpha <adam@outside.com>
Date: Sun, 17 Oct 2004 12:02:39 -0700
MIME-version: 1.0
```

```
Content-type: multipart/mixed; boundary="IronPort"

--IronPort
Content-type: text/plain; format=flowed; charset=us-ascii
Content-transfer-encoding: 7bit

Blah blah blah blah blah
Blah blah blah blah blah
Blah blah blah blah blah
...
--IronPort
Content-type: text/plain
Content-transfer-encoding: 7bit
Content-disposition: inline

X5O!P%@AP[4\PZX54(P^)7CC)7}$EICAR-STANDARD-ANTIVIRUS-TEST-
FILE!$H+H*">X5O!P%@AP[4\PZX54(P^)7CC)7}$EICAR-STANDARD-ANTIVIRUS-TEST-FILE!$H+H*

--IronPort--
```

When mbox-formatted files are parsed, it is desirable not to read too much semantics into the "From" line that separates messages. Because many different utilities write mbox files, there is considerable variation in these lines. However, the "From " line can always be used as a message separator line in order to reliably indicate that a new message has started in the mbox file. In all, there are about 20 known formats for the strings after the "From" message separator, which generally makes it very difficult to parse them.

After the "From" line is an email message in RFC 2822 format, with a series of message body headers followed by a blank line followed by additional message body content.

In order to ensure that messages are properly separated, lines that begin with the string "From" are always prepended by a single ">". Various different variants of mbox files handle lines that begin with ">From" differently. In early implementations of applications that wrote mbox files, these lines were not themselves quoted. AsyncOS log files will always prepend a ">" to lines that begin with one or more ">" characters followed by "From".

Here is an example of an mbox file that contains a message that had lines which contain the starting strings "From", ">From" and ">>>>From" in it:

```
From jtrumbo@example1.com Sun Dec 12 12:27:33 2004
X-IronPort-RCPT-TO: trumbo@example1.com
From: jtrumbo@example1.com
To: trumbo@example2.com
Subject: Quote this, if you dare
Date: Sun, 12 Dec 2004 12:28:00 -0700

The following line is just From
>From A From Line

The following line has quoted >From
>>From A >From Line

The following line has many >>>>From
>>>>>From This line has 4 > characters before From

And this is the last line
```

The end of a message in an mbox format file is traditionally signaled by a blank line. However, this is not always present (although AsyncOS does place it there). When an mbox-format file is parsed, you should signal the end of a message either by the start of a new message (delete the blank line if one is present) or by the end of file.

Another variant in the mbox format called for the length of the message to be signaled in a "Content-Length" field within the message header. That format did not use "From" line quoting. AsyncOS does not use this format and does not insert a Content-Length field.