Understand UNIX mbox Format Usage

Contents

Introduction

Pre-requisites and Important Considerations

UNIX mbox Format Overview

Structure of an mbox File

- Example: Single Message in mbox Format

Message Separation and "From" Line Handling

- Example: Quoting in mbox Files

Message End and Parsing

Summary of Key Points

Introduction

This document describes the UNIX mailbox (mbox) format and its application on Cisco Email Security Appliance (ESA).

Pre-requisites and Important Considerations

- Ensure you understand that AsyncOS on the ESA uses UNIX mbox format when archiving messages and in the `log()` action of message filters.
- The mbox format is ASCII-based and not binary.
- There are multiple variants of the mbox format, which can complicate parsing.
- AsyncOS does not use the "Content-Length" header method or insert a `Content-Length` field.

UNIX mbox Format Overview

UNIX mbox format is used by AsyncOS when messages are archived or logged through message filter actions. "Archive Message" is an additional configuration option for features such as IronPort Anti-Spam (IPAS), Anti-Virus (Sophos and McAfee), Advanced Malware Protection (AMP), and Graymail on the ESA.

The mbox format is a plain-text (ASCII) format that stores one or more email messages concatenated in a single file. Messages are separated using a specific pattern, making it possible to extract individual messages. The format matches messages as transferred between RFC 2821-compliant mail gateways.

Structure of an mbox File

- **Message Separator:** Each message starts with a line beginning with "From " (the ASCII characters F, r, o, m, and a space).
- **Fields After "From ":** The next fields are:
  - **Envelope-sender:** Can be a real mailbox, a dash (-), or another string, depending on the application generating the mbox.
  - **Date:** Inserted by ESA in standard UNIX asctime() format, always 24 characters long.
  - **Optional Data:** Some mbox variants add more fields after the date.

- **Field Separation:** These fields are separated by a single space.
- **Email Message Content:** Follows the separator line in standard RFC 2822 format (headers, blank line, body).

Example: Single Message in mbox Format

```
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

--IronPort--
```

Message Separation and "From" Line Handling

- Because many applications generate mbox files, the format of the "From " separator line can vary. However, you can always use "From " as a reliable message boundary.
- There are approximately 20 known variants for fields following the separator, making automated parsing challenging.
- To prevent confusion, lines within the email body that begin with "From" are always prepended with a single ">". AsyncOS log files will prepend a ">" to any line starting with one or more ">" characters followed by "From".

Example: Quoting in mbox Files

```
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
```

Message End and Parsing

- The end of a message is typically marked by a blank line. AsyncOS always includes this line, but not all implementations do.
- When parsing, treat the start of a new "From " separator line or the end of the file as the end of a message. Remove the blank line if present.
- Some variants use a "Content-Length" field to mark message boundaries, but AsyncOS does not use this method.

Summary of Key Points

- AsyncOS uses standard UNIX mbox format for archiving and logging messages.
- The mbox format is ASCII, concatenates messages, and uses "From " lines as separators.
- AsyncOS ensures lines within messages starting with "From" are quoted with ">" to avoid parsing errors.
- Parsing mbox files requires handling numerous variants, especially with message separator lines.
- Do not expect or rely on a "Content-Length" field in AsyncOS-generated mbox files.