Analysing "Received" header keywords

as antispam and whitelisting measure

Introduction Download Operation Updates

Introduction

I was sort of fed up to discover that some of my antispam procmail rules sometime generate false positives for messages coming from my legitimate correspondents (or better, that my legitimate correspondents send badly behaved mail which triggers my antispam rules, but which I could stand to accept, considering who they come from).

So I set up a little device which will be able to whitelist messages coming either from local hosts inside my LAN, or entering from a list of trusted mail exchangers. To determine the mail exchanger, I use the `Received>` keywords in the message header.

Download and installation

Since the above can be of general use I make available, with no guarantee or liability, what I have done according to the GNU GPL.

Please follow these steps for installation :

- Insert in your `.procmailrc`, in the place most appropriate for it, a line calling my rule, e.g. if `$PMDIR` is your procmail rule directory
  `INCLUDERC = $PMDIR/rc.received`
- Get and install in such directory the following file `rc.received`
- Get and install in the same procmail directory the parsing awk file `rcvdparse.awk`
- Optionally get and install in the same procmail directory or in a directory in your path the analysis utility `extractrcvd.csh` and its support awk file `rcvdanal.awk`
- customize the two awk files in the same way (NB **this is mandatory**)
  - Replace the value of string `mylan` with the class C IP address of your LAN.
  - Replace the value of string `mymx1` and `mymx2` with the unqualified hostnames of the MX of your domain (assumed inside your LAN), i.e. the machines authorized to receive directly incoming mail for your domain.
  - Replace the value of string `myhost` with the unqualified hostname of the machine on which you are installing this stuff
- read the next section for usage

Principles of operation

- `rc.received` analyses the mail message header `Received` keywords.
  - the most recent of such keywords is the one recording the end reception on your machine. In turn your machine could have received the mail from
    - itself (localhost, 127.0.0.1)
    - directly from another machine of your LAN, with the exception of the two MX. Both cases will be considered as "local direct"
    - from one of the two MX (requiring further analysis)
    - any other case is treated as "unknown" and is possibly an error
  - if the message came through an MX, there will be two keywords containing the hostname of the MX. One will record the connection between the MX and your own host, and since it has been dealt with above, should be ignored further.
  - the other keyword will contain the connection to the MX from the next host which in turn can be
    - another host on your LAN, which is considered as "local relayed
    - an external host, which should be the MX of SMTP server of the originating domain. We are interested in the IP address of such machine
- The awk scripts quoted above (called by `rc.received` or called manually) return such IP address (or the fixed values `LOCAL DIRECT`, `LOCAL RELAYED` or `UNKNOWN`)
- `rc.received` compares such returned IP address with the content of a whitelist file `mx.analysis.sort` and sets a variable `TRUSTEDMX=yes` if it matches.
- It is your responsibility to provide the file `mx.analysis.sort`, which should list the trusted IPs of the domains with which you have a regular correspondence. One IP per line, followed by an optional comment. It is up to you to sort the file per domain or per IP according to your choice. Include at the front the three mandatory lines.
  This is an example of `mx.analysis.sort`

```
LOCAL.DIRECT    # do not remove
LOCAL.RELAYED   # do not remove
LOCAL.PERHAPS   # do not remove (same host ?)
140.211.11.2    # spamassassin mailing list on mail.apache.org
```

...

- You can then use `TRUSTEDMX` in further procmail rules as a favourable score to avoid false positives in spam scoring rules. E.g. via conditions like this

```
* -999^0  TRUSTEDMX ?? yes
```

- The (optional) analysis utility `extractrcvd.csh` is supplied as an aid in determining the MX IP of a message (so that you can manually insert it in the whitelist).
  This utility is used piping into it the entire mail message with full headers. The way to do this are dependent on your Mail User Agent. The following works with pine or alpine
  - while viewing a message (or while positioned on it in the message index)
  - issue command **H** to make sure full header display is on
  - issue the pipe **|** command
  - In response to the "Pipe message n to:" prompt, type in the full path of `extractrcvd.csh`
- in reply you will get a listing like this (type **E** to exit the viewer when done)

```
message  From: "Some Body" <Q.U.Alcuno@rcm.inet.it>
Received: from oort.lambrate.inaf.it (oort.lambrate.inaf.it [155.253.16.49])by poseidon.lambrate.ina...
ENTERING VIA oort
Received: from alpha.retecivica.milano.it (alpha.retecivica.milano.it [212.239.120.182])by oort.lamb...
COMING FROM 212.239.120.182
```

  The listing usually includes the From: line, an excerpt of the most recent Received: line, the identification of the MX it is entering through, an excerpt of the relevant Received: line, and finally the IP of the external MX. This value (in red in the example) is what you are interested in.
  For local messages the output is different, but these should be uninteresting at this stage.

Update history

S/w V1.0 and web page established on 19 Jan 2009
Web page last updated on 19 Jan 09 16:20

Bugs

None known yet ... ... but there is no guarantee the analysis will work in any setup (this assumes a setup like hours, where all local machines are on the same LAN, and can receive e-mail from outside only via two MXs which are also on the same LAN)

Contacts

I have no time to actively mantain this software other than for my own use. However I will be glad to receive communication of problems (or improvements) to it at my e-mail address `lucio` in domain `lambrate.inaf.it`.

Lucio Chiappetti - IASF Milano - INAF