SMTP Information gathering

Lluis Mora, Neutralbit
llmora@neutralbit.com

Black Hat Europe
Amsterdam, NL // March 2007
Introduction

• E-mail is present in nearly every organization

• We all understand how it works
  – How envelope and headers work
  – How it can be spoofed
  – How it can be read in transit
  – What a message looks like
  – What to say and what to keep to ourselves

• But what does a message tell about its sender?
What makes SMTP messages so interesting?

Control information is embedded in the message
- Some headers are mandatory, others can be stripped
- All of them usually end up stored in the mailbox

Mailing list archives
- Public logs of our communications
- Stored over the years
- The ultimate SMTP information gatherer source!
• Received headers: an advanced “record route”

  – Probably the most well-known information gathering aspect of SMTP

  – Mandatory, per RFC2821: each node adds its header, no one touches the headers

  – Used to prevent mail loops and debug delivery

  – Strip with caution
Each relay adds
- IP address of sending gateway
- FQDN of receiving server
- Transfer protocol
- MTA server software
- Timestamp, including time zone
SMTP Network mapping (III)

• Not a traceroute…
  – SMTP path, not at the IP level

• … but has its own advantages
  – Allows us to peek behind NAT and firewalls
  – Point-to-point relaying
  – It is initiated by the victim, part of the communication

• Not rocket science
  – Everybody knows about them, but are we conscious of what they tell about us?
Corporate IP subnetting
- Received header addresses are not translated
- Internal IP addressing scheme
- Type of connection to the internet
Corporate Internet access policies

- Centralized Internet access?
- Each location has a public connection?

Received: from mx1.uk.example.com ([195.166.192.8])
  by vger.kernel.org

From John Doe <jdoe@uk.example.com>

Received: from smtp.de.example.com ([32.1.120.11])
  by vger.kernel.org

From Pam Plinas <pplinas@de.example.com>
SMTP Network mapping (VI)

- Server fingerprinting
  - Software and versions
  - Location based on time zones
SMTP Network mapping (VI)

- Relay link information
  - SMTP Link encryption

Received: from lappy (192.168.1.4) by pub.example.net (qmail) with ESMTP
  ID MG0007DA (SSL/TLS, 3DES, CBC mode, keysize 192 bits); 8 Sep 2006 16:40:03 +0200

Received: from [24.26.7.196] (ilm.example.com [24.26.7.196])
  (using TLSv1 with cipher DHE-RSA-AES256-SHA (256/256 bits))
  (No client certificate requested)
• Graphic representation of SMTP paths
  – Definitely flashier than staring at logs
  – Parsing of “Received” headers is challenging
  – Absorb more information at once
  – One image…

• A few examples
  – Data extracted from Linux kernel mailing list
  – Around 3 months in early 2006
spot the telecommuters ...
SMTP Network mapping (VII)

... target selection?
where is wally?

SMTP Information gathering
Client fingerprinting

- Based on a different set of headers
  - User-Agent
  - X-Mailer
  - X-MIME-OLE

- Excellent level of details
  - Down to the patch level

- Not used for anything else
Client fingerprinting (II)

X-Mailer: Microsoft Office Outlook, Build 11.0.5510

User-Agent: Thunderbird 1.5.0.7 (Windows/20060909)

X-Mailer: ColdFusion MX Application Server

X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.2962

X-Mailer: Evolution 2.2.3 (2.2.3-4.fc4)

X-Mailer: iPlanet Messenger Express 5.2 Patch 2 (built Jul 14 2004)

X-Mailer: Lotus Notes Release 5.0.6a January 17, 2001

User-Agent: Squirrel Mail/1.4.3a

User-Agent: Wanderlust/2.12.0 (Your Wildest Dreams) SEM/1.14.6 (Maruoka)
FLIM/1.14.7 APEL/10.6 MULE XEmacs/21.5 (beta21)
(corn) (+CVS-20050720) (i386-suse-linux)
• Long term analysis
  – If we get access to a long stretch of messages
  – Plot client mailers over time…
  – … then add mailer release dates
• Organization trend analysis
  – With enough e-mails, we can find out details about the organization policies
  – Patching policies
  – Application usage
  – Security gaps
  – Policy exceptions

…maybe not just for SMTP servers?
Usage trends

Other interesting facts can be guessed

- Same e-mail address + alternating mailers + multiple IP addresses → multiple locations (home / work?)

- Same e-mail address + same mailer + multiple IP addresses → take the laptop home

- Various e-mail domains + same mailer + same IP address → non-corporate mail at work

- Changing “Date” time zones → user on the go?
Other interesting headers

- Indirect sources of information
  - Implementation differences
    - Ordering of headers
    - Quoted replies
  - Custom X-Headers
    - X-Originating-IP, etc.
    - Antivirus / Antispam

- Message contents
  - User data
  - Encoding data
Other interesting headers (II)

• Indirect sources of information
  – Encoded data in unsuspecting headers

```plaintext
Message-ID: <Pine.LNX.4.21.0611280421440.26304-100000@example.org>

Message-ID: <1103.203.41.53.196.1128283359.squirrel@mail.example.com>

Message-ID: <11363603.1154544476739.JavaMail.root@as.example.net>

Content-Type: multipart/mixed; boundary=Apple-Mail-1-944594902
```
Conclusions

- Strip unneeded information at border gateways whenever possible

- Find out what has already leaked and fix it

- Analysis relies on client provided data, handle with care
Thank you!

Lluis Mora
llmora@neutralbit.com

neutralbit security innovation