Stopping Outgoing Spam by Examining Incoming Server Logs

Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

CEAS, Stanford
22nd July 2005
Summary

• ISP email handling
• Log processing for customers
• Log processing for non-customers
• So how effective are we being?
ISP Email Handling

- Smarthost
- MX host
- The Internet
Heuristics

• Simple heuristics work really well
  – just as for smarthost
• Look for multiple HELO lines
  – often match MAIL FROM (to mislead)
  – may match RCPT TO (? authenticator ?)
• Look for outgoing email to the Internet
• ?? Look at Brightmail’s opinion ??
  – but need to discount forwarding
This report relates to [192.0.2.1] = example.demon.co.uk.

which IP address was in use between 2005-07-20 18:35:27
and 2005-07-20 18:43:39

and from whence an overall total of 4 messages and 0 bounces were sent
these were to 4 destinations of which 0 were failures
and for which we generated 0 reports about undeliverable messages.
The total size of all these messages was 120 KB.

The HELO text is varying so much that relaying or a virus is suspected

There are 4 items in this category

HOST = , HELO = example.demon.co.uk

2005-07-20 18:35:27 first.last@gmx.at -> wpb@example.demon.co.uk Size=30616

HOST = , HELO = other.demon.co.uk

2005-07-20 18:37:09 sales@digitaldepot.co.uk -> clive@other.demon.co.uk Size=30589
2005-07-20 18:37:33 gweek.inn@tesco.net -> nne@other.demon.co.uk Size=30385

HOST = , HELO = demon.net

2005-07-20 18:43:39 sales@example.co.uk -> helpdesk@demon.net Size=30561
Excellent Results

• Four weeks of data from Demon Internet
• Spam source (relay, SOCKS, trojan &c)
  – 78 valid reports
  – 6 false positives, 52 examples missed
• Virus infected
  – 318 valid reports
  – 5 false positives, 88 examples missed
• Low volumes were main reason for errors
The Rest of The Internet

- Can use same heuristics to look at incoming email from the rest of the planet
- Looked at data for just a single day 😞
- 6.6 million emails from 413,728 IP addresses
- 2,527 were virus infected
- 35,615 were sources of spam
- Looked up the AS (ISP responsible)
Viruses

• Figures in the paper
  – #1 was “BTnet” (large UK ISP)
  – #2 was “CHINANET”
  – #3 was “NTL” (large UK ISP)
  – #4 was “Telewest” (large UK ISP)
  – #5 was “Telefonica” (large Spanish ISP)

• Viruses leverage address book contents…
Spam

• Figures in the paper
  – #1 was “CHINANET”
  – #2 was “Korea Telecom”
  – #3 was “China Telecom”
  – #4 was “Hanaro Telecom” (KR)
  – #5 was “AT & T”

• Spammers operate with global lists…
How Many at Demon Internet?

• Clearly number of senders vary depending on size of external network
• BUT percentage infected varies widely
• However, consider the detection method, and note that there’s nothing special about Demon customers – so examine ratio of detection to those sending any email at all…
Detection Ratios

- Asian networks
  - 1 sender in 30..100 detected as “bad”

- European networks
  - 1 sender in 300 detected as “bad”

- Demon Internet in Europe… so maybe we have only one bad customer in 300 as well?
Demon Internet

• On relevant day, spotted 42 Demon customers with problems
• Hence 42 x 300 (12,000) actually have problems (perhaps?)
• System picked up 530 over the month (including the false negatives)
• Only 8445 used the incoming servers at all!
Conclusions

• Processing incoming server logs works
• You can learn a lot about where spam comes from (and who is virus infected)
• Figures suggest that only picking up around 5% of ISP’s problem – which is a start (so don’t knock it), but not especially cheering
• More log processing ideas next year 😊
Stopping Outgoing Spam by Examining Incoming Server Logs

Richard Clayton

http://www.cl.cam.ac.uk/~rnc1/incoming.pdf