spamHINTS update

PI: Prof. Ross Anderson
Researcher: Dr. Richard Clayton
Happily It’s Not The Same

• The sending of spam differs from the sending of legitimate email, not just in content but in the traffic patterns
• *Time* email is “9 to 5”, spam is 24 hours
• *Space* spam goes to many destinations or all to just one ISP (in a “dictionary” attack)
• *Size* spam is a constant size
• Virus/Worm traffic is like spam (but bigger)
Summary

Essentially a (limited) progress report

- Funding
- Preparation
- sFlow monitoring
- Email server log processing
  - Internal (datamine your logs to spot abuse)
  - External (pass reports of abuse to others)
  - Best Practice Document
Funding situation

• Intel Research
  – will do second year if satisfactory progress

• NTL
  – non-committal commitment

• Department of Industry
  – poor & wanted to see industry funding first

• LINX
  – data processing, websites etc
spamHINTS @ LINX
Preparation

• Processing route tables (RIS etc) tells you which AS owns address space
  – except when there is overlap or error 😞

• Processing RIR databases tells you contact addresses for AS’s
  – registries protective of this data
  – data is unstructured and incomplete 😞
sFlow data processing

• Delayed by other commitments 😞
  – employed on spamHINTS since 1 Feb
• Have developed (with LINX) short term plans for capturing some example data
  – want one day’s worth for initial analysis
• One minute’s worth of data shows that sFlow also contains content (!!!)
  – submitted patches for sflowtool to fix this
ISP email handling

Smarthost

The Internet
Log processing: #1 internal

• Want to encourage more ISPs to process their server logs
  – proven technique
  – saves you time/money/blacklisting

• De-demon-ising log processing Perl is taking longer than expected
  – Real Soon Now!
Log processing: #2 external

• Want to encourage ISPs to share email log info about incoming spam & viruses
• Send report to host ISP indicating:
  – source IP address (and of course time)
  – source email address (probably forged)
  – destination email address
  – metadata (size, HELO message, filter results)
  – diagnosis of problem
• Reporting is straightforward except…
• … email addresses are personal data
  – Information Commissioner quite clear on this
• Much is of course forged, but amongst this
  may be some real email, and source/destination details could be sensitive
  – so must meet legal obligations
Legitimate processing

• Asking another ISP to take action to prevent their user sending spam/virus traffic can be seen as legitimate processing

• So jump through correct hoops & all OK
  – inform customers of processing
  – (try to) inform senders of processing
  – ensure processing covered by privacy policy
  – address any promises of confidentiality
Best Practice document

• Would also be desirable for processing to be in line with industry Best Practice!

• Hence recent draft of:

   Best Practice for reporting abuse issues based on traffic data
Components of Best Practice

• Reports based on traffic data must only be sent by prior agreement
• Reports should not be unduly repetitive
• The evidence on which the report is based must be clearly given (& accurately timed)
• Needs warning about personal data
• Must keep customers informed (as above)
Outstanding Best Practice issues

• Outside of EU raises other problems
  – can probably resolve via contract, but may make it just too much trouble 😞

• Some unresolved comments
  – Davies: Best Practice to report statistics?
  – Cormack: Encourage special email address?

• …any more comments today ?
Richard Clayton
<rnc1@cl.cam.ac.uk>

www.spamhints.org