Modelling Incentives for Email Blocking Strategies

Andrei Serjantov
Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

WEIS 2005
KSG, Harvard
2nd June 2005
Summary

• Setting the scene
• The model
• The implications of the model
• What is the pattern of outgoing email
• What is the pattern of incoming email
• Where next?
Setting the scene

• Email goes via ISP “smarthosts”
• Blacklists identify spam sources
  – may be a factor for Bayesian classifiers
  – may be used to block the sender altogether
• ISPs act in an ad hoc manner doing what seems to make sense to their sysadmins, and sometimes their customers
• Blacklists pretty much ad hoc as well!
The Model

A

B

C
The Model

- Utility of ISP depends on its connectivity
  - Positive: ability to send email to others
    - Depends on how many people there are “out there”
  - Positive: reception of good email from others
    - Hard to perceive (all sorts of possible errors): ignore this term
  - Negative: reception of spam from others
    - Depends on how vulnerable remote clients are
    - And how many clients we have they may send to

\[
Utility (A) = \left( \sum_B U(C_B) \right) - \left( |C_A| \times \sum_B V_B(C_B) \right)
\]
Implications of the model

• The more “vulnerable” your clients are the bigger the negative term other ISPs see
  – they have to estimate this: guard your reputation!
• Dictionary attack spam affects large ISPs more (they have more clients who see it)
• Tit-for-tat blocking may work: remote ISP blocking us, we block them, our users don’t notice (!) but their users do
The view from large ISPs

- To large ISPs rest of world is very small
- Hence utility of connection to remote ISP dominated by how much spam they send
- Furthermore, utility equation dominated by self-sending term, and hence internal controls should be the overriding concern!

$$Utility\ (A)_{self} = U(C_A) - |C_A| \times V_A(C_A)$$
Outgoing email

• Measured outgoing email from Demon Internet (medium sized UK ISP) for four week period in March
  • excluded virus infected, spam sources etc
  • 82 000 customers (>50% use Hotmail etc)
  • 25 245 000 emails (of which 9 857 000 “bounces”)
  • 378 821 destination MX servers
  • but 240 850 only used once (typos + spam rejects)
<table>
<thead>
<tr>
<th>ISP</th>
<th>Rôle</th>
<th>emails</th>
<th>customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>msgelabs.com</td>
<td>Spam filtering</td>
<td>1,361,916</td>
<td>35,641</td>
</tr>
<tr>
<td>hotmail.com</td>
<td>Global webmail</td>
<td>1,320,900</td>
<td>43,350</td>
</tr>
<tr>
<td>aol.com</td>
<td>Global ISP</td>
<td>820,645</td>
<td>37,674</td>
</tr>
<tr>
<td>btinternet.com</td>
<td>UK ISP</td>
<td>809,367</td>
<td>39,048</td>
</tr>
<tr>
<td>yahoo.co.uk</td>
<td>UK portal</td>
<td>367,327</td>
<td>24,302</td>
</tr>
<tr>
<td>demon.net</td>
<td>UK ISP (self)</td>
<td>363,112</td>
<td>15,212</td>
</tr>
<tr>
<td>ntl.com</td>
<td>UK ISP</td>
<td>337,441</td>
<td>25,174</td>
</tr>
<tr>
<td>yahoo.com</td>
<td>Global portal</td>
<td>298,491</td>
<td>18,139</td>
</tr>
<tr>
<td>uk.tiscali.com</td>
<td>UK ISP</td>
<td>235,858</td>
<td>22,022</td>
</tr>
<tr>
<td>virgin.net</td>
<td>UK ISP</td>
<td>189,389</td>
<td>18,358</td>
</tr>
<tr>
<td>schlund+partner</td>
<td>German web hosting</td>
<td>166,077</td>
<td>14,540</td>
</tr>
<tr>
<td>nhs.uk</td>
<td>UK health service</td>
<td>160,793</td>
<td>9,816</td>
</tr>
<tr>
<td>blueyonder.co.uk</td>
<td>UK ISP</td>
<td>149,521</td>
<td>14,677</td>
</tr>
<tr>
<td>pipex.net</td>
<td>UK ISP</td>
<td>97,495</td>
<td>9,576</td>
</tr>
<tr>
<td>Spicerhaart.co.uk</td>
<td>UK estate agent</td>
<td>85,425</td>
<td>144</td>
</tr>
<tr>
<td>clara.net</td>
<td>UK ISP</td>
<td>82,309</td>
<td>8,106</td>
</tr>
<tr>
<td>mailcontrol.com</td>
<td>Spam filtering</td>
<td>80,941</td>
<td>6,978</td>
</tr>
<tr>
<td>global.net.uk</td>
<td>UK ISP</td>
<td>77,957</td>
<td>10,586</td>
</tr>
<tr>
<td>plus.net</td>
<td>UK ISP</td>
<td>77,080</td>
<td>9,289</td>
</tr>
<tr>
<td>postini.com</td>
<td>Spam filtering</td>
<td>74,777</td>
<td>6,092</td>
</tr>
</tbody>
</table>
Destinations: amount of email

• Power law distribution
  – see paper for straight line graph
• viz: same amount of email being sent to top 10 sites as to the next 100 as to the next 1000 as to the next 10000…
• A strategy that keeps only 10 destinations sweet (or only 100 etc) will fail
Destinations: number of senders

13 sites >10,000 customers sending to them
213 sites >1,000 customers sending to them
2601 sites >100 customers sending to them

• Potential for many complaints if just one of many other ISPs blocks Demon’s email
• How much should Demon spend on their abuse team?
  – clearly has a simple answer: Enough!
Incoming email

- 14 days incoming email
- 55.6 million emails
- 66.5% categorised as spam by “Brightmail”
- 13,378 sending ASs
- If an AS sent nothing but spam then would be rational to bar them
  - early test: one AS sent 9948, all spam in a day
Incoming: results inconclusive

- Many sources sent mainly spam, but still a few a day that were not
- Large volumes of spam (which would make real difference) accompanied by large volumes of good email
- Much more study needed
  - results much influenced by Brightmail
  - fast responses needed (infamous AS now OKish)
Conclusions

- Model explains much real world behaviour
- Figures clearly show very diverse aspect to communications: so ISPs cannot operate on a handful of special relationships
- Barring incoming email without impacting real traffic doesn’t look simple
- Still believe rational strategies are possible
Modelling Incentives for Email Blocking Strategies

Andrei Serjantov
Richard Clayton

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