Evaluating The Use of Spam-triggered TCP/IP Rate Control To Protect SMTP Servers

Minh N. Tran
Grenville J. Armitage

Talk outline

- What is Spam?
- Our anti-spam tool: MT Proxy
- A basic email architecture
- Traditional and new anti-spam techniques
- Design and functionality of MT Proxy
- Experimental validation
- Research limitation and future work
- Conclusion
What is Spam?

- “Mass unsolicited electronic mail”
  (RFC 2505, “Anti-Spam Recommendations for SMTP MTAs”,
  http://www.ietf.org/rfc/rfc2505.txt)
- “Unsolicited Bulk Email ("UBE")”
  - Unsolicited: “Recipient has not granted verifiable permission for the message to be sent”
  - Bulk: “Message is sent as part of a larger collection of messages, all having substantively identical content”
  (The Spamhause project, “Definition of spam”,
  http://www.spamhaus.org/definition.html)
- “Unsolicited e-mail, often of a commercial nature, sent indiscriminately to multiple mailing lists, individuals, or newsgroups: junk e-mail”

Talk outline

- What is Spam?
- Our anti-spam tool: MT Proxy
- A basic email architecture
- Traditional and new anti-spam techniques
- Design and functionality of MT Proxy
- Experimental validation
- Research limitation and future work
- Conclusion
Our anti-spam tool: MT Proxy

- A new anti-spam approach
- Inspired by Marty Lamb’s exhortation
  - “We want to cause spammer pain”
- Two goals of MT Proxy:
  - To cause resource consumption at spammer end
  - To avoid negative impact of legitimate emails being misclassified as spam

Talk outline

- What is Spam?
- Our anti-spam tool: MT Proxy
- A basic email architecture
- Traditional and new anti-spam techniques
- Design and functionality of MT Proxy
- Experimental validation
- Research limitation and future work
- Conclusion
A basic model of email transfer

- Simple Mail Transfer Protocol (SMTP, defined in RFC 821)

![Email Transfer Diagram]

Role of MT Proxy

- MT Proxy sits in front of recipient’s MTA to intercept SMTP traffic coming to this server

![MT Proxy Diagram]
Talk outline

- What is Spam?
- Our anti-spam tool: MT Proxy
- A basic email architecture
- Traditional and new anti-spam techniques
- Design and functionality of MT Proxy
- Experimental validation
- Research limitation and future work
- Conclusion

Traditional anti-spam techniques

- White and black listing:
  - Focus on addresses associated with the email
  - Emails in White list are legitimate
  - Emails in Black list are spam
  - Challenge response mechanism

- Rule-based filtering:
  - Inspect the actual content of the email
  - Bayesian algorithm: most spam events are dependent
Challenges for traditional methods

- Trade-off between False Positive and False Negative
  - False Positive: A legitimate email is incorrectly identified as spam
  - False Negative: A spam email is incorrectly classified as non-spam
- ISPs implement more aggressive anti-spam methods
  - Loss of over $50 per person per year and $3.5 billions per a U.S. business in 2003 due to false positive (Research by Ferris Inc.)
- Need to eliminate false positives
- Traditional methods leave no painful impact on spammer

New anti-spam approaches

- Sender Policy Framework (SPF)
  - Domain authentication technique to identify spam forgery
  - Maintain registered domain names and their associated mail servers
- Anti-spam router (ASR) of TurnTide
  - Allocate different Quality of Service (QoS) for different incoming email traffic according to its spam level
- Microsoft’s “stamp of approval”
  - Delay is added to SMTP traffic through cryptographic puzzles solved by sender
- MT Proxy
  - Eliminate the negative consequences of false positives
  - Shift back the cost to spammers
Talk outline

• What is Spam?
• Our anti-spam tool: MT Proxy
• A basic email architecture
• Traditional and new anti-spam techniques
• Design and functionality of MT Proxy
• Experimental validation
• Research limitation and future work
• Conclusion

Design and functionality

• FreeBSD proxy server intercepting email traffic
• Use blacklisting and content filtering
• Blacklisting: local and Internet DNS server
• Content filtering: spammer experiences slower connection in real-time
• Traffic is shaped at TCP/IP level using FreeBSD kernel’s resident: ipfw and dummynet
Talk outline

• What is Spam?
• Our anti-spam tool: MT Proxy
• A basic email architecture
• Traditional and new anti-spam techniques
• Design and functionality of MT Proxy
• Experimental validation
• Research limitation and future work
• Conclusion

Experimental validation

Testbed for evaluating the efficacy of MT Proxy
Experimental validation: Spam vs non-spam

- Clients use open source software smtpclient to send 500 non-spam, spam, non-spam/spam emails to user1, user2, user3 respectively
- Each client send emails as fast as possible, but do not overlap itself
- From the graph, MT Proxy has proved to be capable of slowing down spam

Experimental validation: Email size

- When email size increases, the relative time delay between spam and non-spam also increases
- When email size increases, the mail box size reduction also improves

<table>
<thead>
<tr>
<th>Email size (in MB)</th>
<th>Average delay for spam email (in sec)</th>
<th>Number of non-spam emails received per second</th>
<th>Number of spam emails received per second</th>
<th>Spam mail box size reduction per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>6.59</td>
<td>5.45</td>
<td>342</td>
</tr>
<tr>
<td>2</td>
<td>6.4</td>
<td>4.04</td>
<td>1.32</td>
<td>1665</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>2.63</td>
<td>2.05</td>
<td>2812</td>
</tr>
<tr>
<td>10</td>
<td>120</td>
<td>2.59</td>
<td>2.04</td>
<td>4720</td>
</tr>
<tr>
<td>20</td>
<td>240</td>
<td>2.40</td>
<td>1.74</td>
<td>8752</td>
</tr>
<tr>
<td>50</td>
<td>450</td>
<td>2.11</td>
<td>1.80</td>
<td>52750</td>
</tr>
<tr>
<td>100</td>
<td>5000</td>
<td>1.57</td>
<td>0.13</td>
<td>341*456</td>
</tr>
<tr>
<td>200</td>
<td>24000</td>
<td>1.03</td>
<td>0.04</td>
<td>194004</td>
</tr>
</tbody>
</table>

Spam reduction for different email sizes (100% spam case)
Experimental validation: Email size

- After the threshold, spam emails are received at a significant time after non-spam emails

![Graph: Time delay as a function of email size]

Experimental validation: Spam structure

- 4 types of email structure in which the top part is non-spam and the bottom part is spam
- The earlier spam appears in the email, the more effective MT Proxy is

![Graph: Average time delay as a function of email size for 4 types]
Talk outline

- What is Spam?
- Our anti-spam tool: MT Proxy
- A basic email architecture
- Traditional and new anti-spam techniques
- Design and functionality of MT Proxy
- Experimental validation
- Research limitation and future work
- Conclusion

Research limitation and future work

- The architecture of MT Proxy does not work very effectively for spam emails with small size
  - Email size ranges from 1Kbytes to 11Kbytes with mean of 4.64 Kbytes (6955 spam emails logged by our University’s IT department on 2 Aug 2004)
  - MT Proxy is more effective against email bodies where signs of spam occur early
  - Improved version of MT Proxy learns which source IP addresses had attempted to send spam to rate-limit their subsequent connection
  - Sending many small emails will be equivalent to sending a single large consecutive email
Talk outline

• What is Spam?
• Our anti-spam tool: MT Proxy
• A basic email architecture
• Traditional and new anti-spam techniques
• Design and functionality of MT Proxy
• Experimental validation
• Research limitation and future work
• Conclusion

Conclusion

• MT Proxy sits in front of recipient’s SMTP server
• MT Proxy has proved to have certain contribution to the arsenal of anti-spam techniques
• MT Proxy can effectively slow down traffic from spammers using FreeBSD’s kernel resident ipfw/dummynet
• MT Proxy’s approach also avoids the damaging consequences of false positives because all email eventually gets through
• We are developing an improved version of MT Proxy with long term memory of who is sending spam