Improving Efficiency of E-mail Communication via SPAM elimination using blacklisting

Tomáš Sochor, Radim Farana
Department of Informatics and Computers
University of Ostrava
Czech Republic

21st TELecommunications FORum
Beograd November 2013
What SPAM is and why is it risky?

- SPAM in the sense of Unsolicited e-mail message
  - originated in 1970’s
  - almost omnipresent NOW
    - >80~90% of all e-mail messages are SPAM
- Main risks of SPAM:
  - Vector for viruses and malware spreading
  - SPAM consumes majority of server resources
SPAM ratio trends
global & local (SPAM No)/Total
anti-SPAM means should be applied HERE

- Server under control
- Before distribution to users

21st TELEcommunications FORum
Beograd November 2013
Typical SPAM control system with feedback

Internet → SMTP server Input Filtering → Greylisting → Virus Scanner

SMTP server

Mailboxes

IP address of SPAMmer to remove from AWL

21st TELecommunications FORum
Beograd November 2013
Several ways of SPAM elimination

- Blocking messages before delivery
  - Formal blocking
    - E.g. for non-existent recipient address
  - Blacklisting
  - Greylisting
    - Temporary refusing from unknown sources
- Detection after delivery
  - Simplest and oldest methods
    - Seldom efficient
Greylisting

- Efficient tool to stop SPAM
  - filtering SPAM before delivery
    - saves server disk space and other resources
- Greylisting use
  - Proposed in 2003 (E. Harris)
  - 2006 – University of Ostrava
  - Widely used for now
    - Still very efficient
Greylisting principle

- Basic principle is simple:
  - operates BEFORE message delivery
  - inserting short delay in message delivery
    - approx. 5 – 10 minutes
  - SPAMmer does not repeat the attempt
    - except very sophisticated ones - rare
  - applied only to unknown sources
    - to minimize the impact to the service
aaa.cz
SMTP sender

220 bbb.cz SMTP Server Ready
HELO aaa.cz
250 bbb.cz Hello aaa.cz
MAIL From: sender@aaa.cz
250 Sender OK
RCPT To: recipient@bbb.cz
250 Recipient OK
DATA
354 Enter mail ...
e-mail
e-mail ...
.
250 Mail accepted
BYE
221 Closing connection

bbb.cz
SMTP recipient

Error message responded here

21st TELecommunications FORum
Beograd November 2013
Greylisting weakness

- It is easy to adapt to greylisted server
  - so far it seems not efficient for SPAMmers
- SPAMmer can get into AWL:
  - After several successful deliveries through greylisting it is considered to be a legal source
    - and not checked any more
  - this behaviour can be eliminated by connection with SPAM scanner
    - also DNSBL in the future?

21st TELEcommunications FORum
Beograd November 2013
Blacklisting

- Blocking of messages from certain sources
  - Problem: sender e-mail address can be falsified
    - Easily rewritten in the message
  - IP address of the sending SMTP server are used
- Serious issue:
  - How to keep the blacklist up-to-date
    - Usually 3rd party blacklists are used
Blacklisting results

- It works well
  - Blocks significant part of SPAM messages
- Its efficiency is roughly comparable to greylisting
  - However 3rd party blacklist is used
    - Spamhaus.org

21st TELcommunications FORum
Beograd November 2013
Efficiency of SPAM elimination
Blacklisting and Greylisting
Blacklisting requests
Potential errors of Blacklisting

Later check performed in March 2013
Blacklisting Issues

- 3rd party dependency
  - Blacklist source
- False positive errors are dangerous
  - Result in refusal the delivery of legitimate message
  - Such errors were traced from logs
    - Not automatically
    - The search was limited
      - to selected periods
      - To senders from selected Top Level Domains
        - The most frequently used (.cz, .sk, and 4 more)

21st TELecommunications FORum
Beograd November 2013
Blacklisting Errors

- Error detection
  1. Selection of blocked senders from .cz, .sk, .at, .pl, .de
  2. Removing addresses with unexpected characters
  3. Verification of continuous blacklisting of the address
  4. Approximation of the geoposition of the sender (non-corresponding removed)
  5. Address verification
     - verify-email.org
## Blacklisting Issues

<table>
<thead>
<tr>
<th>Period</th>
<th>Err1</th>
<th>Err2</th>
<th>Total Requests</th>
<th>Error Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2009</td>
<td>1</td>
<td>4</td>
<td>45,478</td>
<td>0.01%</td>
</tr>
<tr>
<td>December 2010</td>
<td>12</td>
<td>6</td>
<td>47,015</td>
<td>0.04%</td>
</tr>
<tr>
<td>December 2011</td>
<td>15</td>
<td>1</td>
<td>47,084</td>
<td>0.03%</td>
</tr>
<tr>
<td>December 2012</td>
<td>5</td>
<td>1</td>
<td>44,530</td>
<td>0.13%</td>
</tr>
<tr>
<td>January 2013</td>
<td>2</td>
<td>9</td>
<td>2,836</td>
<td>0.39%</td>
</tr>
<tr>
<td>February 2013</td>
<td>4</td>
<td>1</td>
<td>3,148</td>
<td>0.16%</td>
</tr>
</tbody>
</table>

Err1.. Newsletters, Err2 ... ordinary messages
SPAM Elimination Efficiency
greylisting vs. content scanner
Conclusions – Main observations

- Blacklisting is efficient
  - Elimination of 50 – 75% SPAM messages
- Blacklisting is not error-free
  - But the error ratio is very low
- best when combined with other techniques, e.g.
  - Greylisting
  - SPAM scanner (content-based)
Questions?
Comments?

- e-mail: tomas.sochor@osu.cz
- home page at the University of Ostrava: http://www1.osu.cz/home/sochor/en/
Greylisting efficiency comparison

- 3 SMTP servers as mentioned
- Short period data available from other sites
  - only single figure
  - average for the period available

<table>
<thead>
<tr>
<th></th>
<th>GL efficiency avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ostrava Uni</td>
<td>89.3%</td>
</tr>
<tr>
<td>Nitra Uni</td>
<td>95.0%</td>
</tr>
<tr>
<td>Zebra</td>
<td>92.2%</td>
</tr>
</tbody>
</table>