Copyright

Copyright Marc Wallman and Theresa Semmens, 2006. This work is the intellectual property of the authors. Permission is granted for this material to be shared for non-commercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the authors. To disseminate otherwise or to republish requires written permission from the author.
Open Source E-mail with Spam Filtering for the Enterprise

North Dakota State University
Fargo, ND
Overview

• Introductions
• About NDSU
• Description of Problem & Solution
• Development & Implementation
• Results
• Final Thoughts
About NDSU

- Land-grant institution – 1890
- 12,000 students, 3000+ faculty and staff
- Part of HECN (Higher Education Computing Network)
- Provides e-mail services for six other ND institutions
Size and Scope of E-mail Services

- Typical message delivered <18 seconds
- Two e-mail systems
  - IMAP
  - GroupWise (approx. 400 accounts)
- Process e-mail for 23,000 addresses
  - Host IMAP e-mail accounts for 18,000
- Accept approx. 135,000 messages daily for approx. 210,000 recipients
Problem

Current spam filter not efficient

- Spam and unwanted e-mail received by campus community
- Diverse campus opinion on filtering e-mail
- 15-20 angry complaints per day received by Help Desk
Solution Steps

• Solution required by University System Policy
• Users threatened lawsuits
• Plan of action
  – Spam Abatement Solutions Group
  – Spam Abatement Technical Group
• Project plan
  – ITLT approval
  – Provost approval
Spam Filter Solution

• An open source interim solution with varying levels of spam filtering the community could choose from
  – Year one budget: $35,440 (.5 FTE + hardware, support)
  – Year two budget: $21,380 (.25 FTE + hardware, support)
  – Year three budget: $15,130 (.25 FTE + hardware, support)

• Implement by August 2005
# Filter Solution Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Obvious Spam</th>
<th>Potential Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disabled</td>
<td>Deliver</td>
<td>Deliver</td>
</tr>
<tr>
<td>1</td>
<td>Quarantine</td>
<td>Deliver</td>
</tr>
<tr>
<td>2</td>
<td>Quarantine</td>
<td>Quarantine</td>
</tr>
<tr>
<td>3</td>
<td>Delete</td>
<td>Quarantine</td>
</tr>
<tr>
<td>4</td>
<td>Delete</td>
<td>Delete</td>
</tr>
</tbody>
</table>
Timeline

- Nov. 1, 2004 to August 2005: R&D
- April 1, 2005: Create policy
- May 1- July 30: Testing phase
- July 1-15: Survey Test Users
- July 30: Final assessment by testing group
  - Services Web page updated to include spam filter
  - Implement Safe & Block Sender Lists
- August 15: Filter is available service
- December 15: Review of project
Method: Tag and Deliver

- Delivery decision based on user selected rules.
- Tag all incoming mail as
  - Not Spam,
  - Potential Spam, or
  - Obvious Spam.

Notes:
- Special tagging to accommodate Novell’s GroupWise.
- Manual delivery rules required for users forwarding their mail.
Benefits of Tag and Deliver

1. Quarantine is managed by user’s e-mail client.
2. Quarantined messages count against user’s quota.
Extra Features

1. Automatic purge of messages > 30 days in quarantine
2. Safe Sender and Block Sender lists
3. Re-registering for spam service will fix a user’s broken spam filter (e.g., a deleted quarantine folder)
Infrastructure

Incoming Mail

Routing decision based on directory information

Mail Router
AV & Spam Tagging

Mail Server
Spam Delivery Decision

Mail Server
Spam Delivery Decision

Mail Server
Spam Delivery Decision

LDAP Directory

Perdition
IMAP Proxy

Connection to mail server based on directory information

IMAP Clients
Advantages

1. Opt-in service
2. Solution integrated with existing services
3. Well defined, easy for end users to understand
4. Tagging method isolates mechanics of spam identification from user interface
5. All mail is tagged: Metrics on spam identification effectiveness are precise
6. Easy infrastructure maintenance
7. Scalable infrastructure
Disadvantages

1. Cannot make delivery decisions for users on systems where spam filtering takes place
2. Cannot make delivery decisions for back end mail servers (e.g., Listserv server) on systems where spam filtering takes place
Subscription Rate

Spam Subscribers by Date

- Disabled
- Level 1 (QD)
- Level 2 (QQ)
- Level 3 (XQ)
- Level 4 (XX)
- Total

Day

Subscribers
Effectiveness
What it looks like
Tools Used

• Sendmail (http://www.sendmail.org)
• MailScanner (http://www.mailscanner.info)
• Spamassassin (http://spamassassin.apache.org)
• Perdition Mail Retrieval Proxy (http://www.vergenet.net/linux/perdition/)
• UW IMAPd (http://www.washington.edu/imap)
• Procmail (http://www.procmail.org)
Questions?

Marc Wallman  
Sr. Systems Administrator  
Marc.Wallman@ndsu.edu  
(701) 231-7168

Theresa Semmens  
NDSU IT Security Officer  
Theresa.Semmens@ndsu.edu  
(701) 231-5870