A Collaborative Approach to Anti-Spam

Chia-Mei Chen
National Sun Yat-Sen University
TWCERT/CC, Taiwan

Agenda

- Introduction
- Proposed Approach
- System Demonstration
- Experiments
- Conclusion

Problems of Spam Mail

- Commercial Spam
  - Reduce productivity
  - waste network bandwidth and increase processing load of mail servers
  - Spam mail may include pornography messages

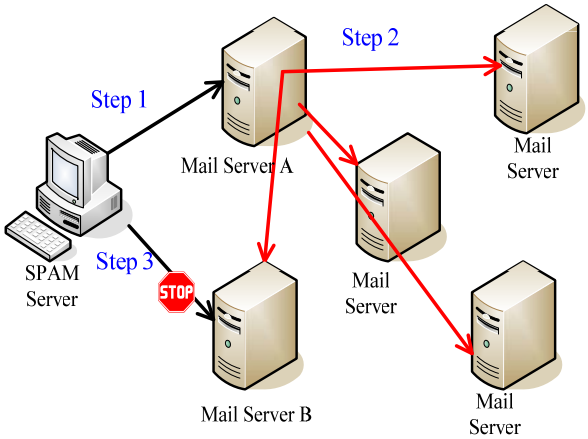
Problems of Spam Mail (2)

- Malicious Spam
  - Virus Spam
  - Worm Spam
  - Rootkit Spam
  - Backdoor Spam
  - Botnet
  - Phishing

Spam Filter

- Most Spam filter is standalone
- Filtering out spam mails based on mail header and keywords
- The most important problem of standalone spam filter is
  - the content of unsolicited messages evolve and may change time by time
  - a standalone mail filter might not be able to fast enough to catch up all new types of spam mails

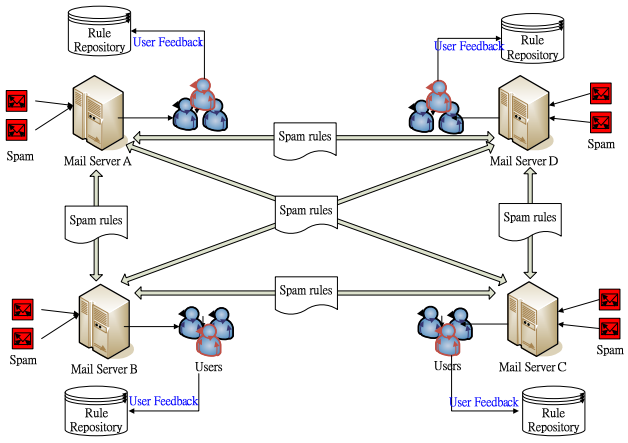
Collaborative Anti-spam Framework



Proposed System

- Spam rule generation
- Spam rule exchange
- Spam rule evolution

System Architecture



Spam Rule Generation

- Using machine learning or statistic approach to generate exchangeable spam rules
  - Decision tree
  - Rough set
  - Bayesian
- Using header information, keyword frequency and format information as feature

Selected Attributes

Attributes	Description
From	The sender's name and email address.
Reply to	If this mail specifies an address for replies to go to
CC	If this mail has carbon copy
Received	It means where the message originated and what route it took to get to you.
Subject	The subject of this mail.
Body	The content of this mail.
Length	The length (byte) of this mail
Domain	The domain name of sender's mail server
Multi part	Does this mail be multi part?
Text/Html	The format of the content of mail.
Hasform	Does this mail have form?
Table	Does the content of mail have tables
Rec_number	The number of keyword in the mail
Encoding	The encoding of this mail

Rule Exchange

```
<rule>
  <if support="7" coverage="0.0187166">
    <and>
      <descriptor attribute="length" value="1"/>
      <descriptor attribute="domain" value="yahoo.com.tw"/>
      <descriptor attribute="ranking" value="0"/>
    </and>
  </if>
  <then>
    <or>
      <decision support="5" accuracy="0.714286" coverage="0.0188679">
        <descriptor attribute="spam" value="1"/>
      </decision>
      <decision support="2" accuracy="0.285714" coverage="0.0183486">
        <descriptor attribute="spam" value="0"/>
      </decision>
    </or>
  </then>
</rule>
```


Spam Rule Evolution

- R_{ii}:the reward
- S_i:the strength of rule i
- S_i can be viewed as rule quality

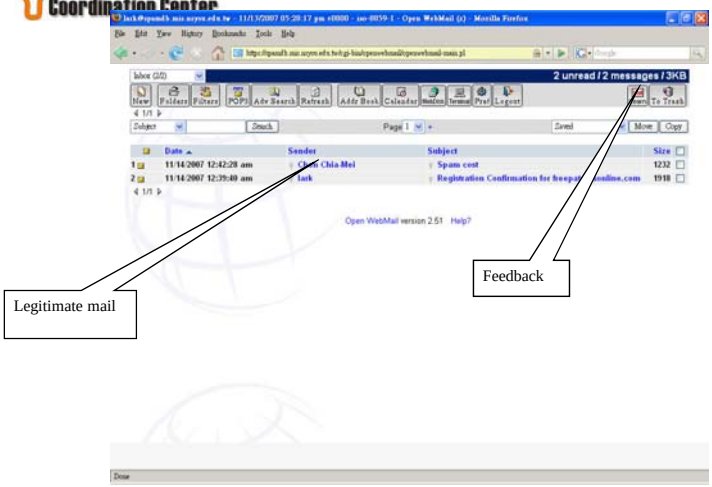
$$S_i = \begin{cases} S_i + R_{ii} & \text{if rule } i \text{ is used} \\ \beta \cdot S_i,\ 0 < \beta < 1 & \text{if rule } i \text{ is not used} \end{cases}$$

$$R_{ii} = \begin{cases} R_{ii} & \text{if classify correctly} \\ -R_{ii} & \text{if classify incorrectly} \end{cases}$$

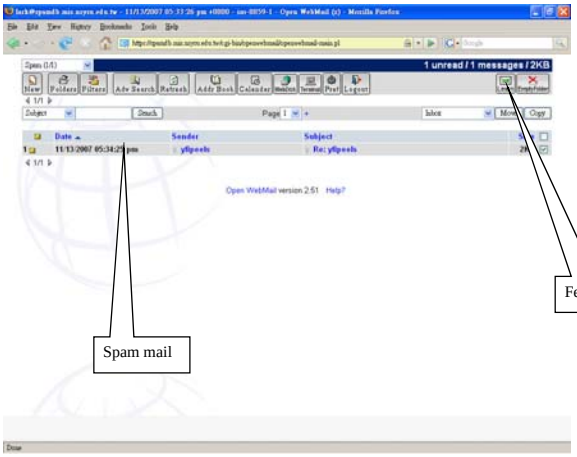
System Demonstration

- User Interface (mail client)
  - Open web mail
- Rule Generate
  - Rosetta
- Mail Pre-Process and Filter
  - Procmail
- Rule Exchange
  - XML Files
- Mail and Rule Repository
  - MySQL Database

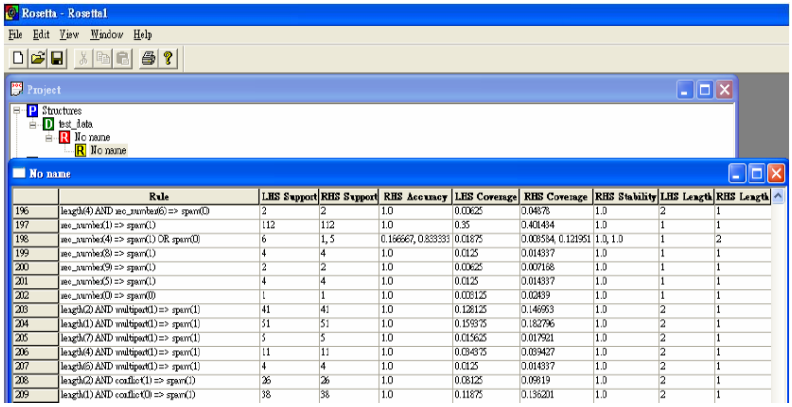
User Interface (Inbox)



User Interface (Spam folder)

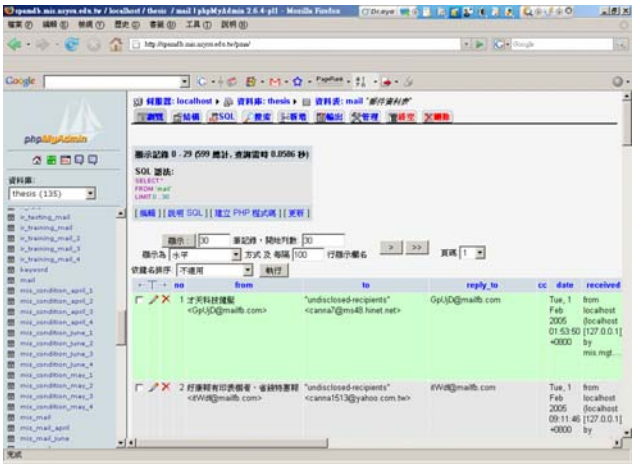


Rule Generation (Rosetta)



	Rule	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage	RHS Stability	LHS Length	RHS Length
196	length(4) AND sto_number(6) => spam(0)	2	2	1.0	0.00625	0.04878	1.0	2	1
197	sto_number(1) => spam(1)	112	112	1.0	0.35	0.401434	1.0	1	1
198	sto_number(4) => spam(1) OR spam(0)	6	1, 5	0.166667, 0.833333	0.01875	0.008584, 0.121951	1.0, 1.0	1	2
199	sto_number(8) => spam(1)	4	4	1.0	0.0125	0.014337	1.0	1	1
200	sto_number(9) => spam(1)	2	2	1.0	0.00625	0.007168	1.0	1	1
201	sto_number(5) => spam(1)	4	4	1.0	0.0125	0.014337	1.0	1	1
202	sto_number(0) => spam(0)	1	1	1.0	0.003125	0.02439	1.0	1	1
203	length(2) AND multipart(1) => spam(1)	41	41	1.0	0.128125	0.146953	1.0	2	1
204	length(1) AND multipart(1) => spam(1)	51	51	1.0	0.159375	0.182796	1.0	2	1
205	length(7) AND multipart(1) => spam(1)	5	5	1.0	0.015625	0.017921	1.0	2	1
206	length(4) AND multipart(1) => spam(1)	11	11	1.0	0.034375	0.039427	1.0	2	1
207	length(6) AND multipart(1) => spam(1)	4	4	1.0	0.0125	0.014337	1.0	2	1
208	length(2) AND conflict(1) => spam(1)	26	26	1.0	0.08125	0.09319	1.0	2	1
209	length(1) AND conflict(0) => spam(1)	38	38	1.0	0.11875	0.136201	1.0	2	1

Mail Repository



Performance Evaluation

- Performance Metrics
- Training and testing data source
- Experiment results

Performance Metrics

- Spam precision
- spam recall
- accuracy
- Miss rate

Data Source

	TWCERT/CC	MIS Department	NSYSU University
Spam mails	3,483	3,115	17,948
Legitimate mails	809	531	991
Totals	4,294	3,646	18,939

Data are gathered from 2006/5/10 to 2006/5/30

Experiment Result- Spam Precision

	10-May	20-May	30-May
Rule A	99.8947368%	98.7804878%	99.2448759%
Rule A ∪Rule B	98.7538491%	98.7912088%	99.3690852%
Rule A ∪Rule B ∪Rule C	98.7551867%	98.7978142%	99.4780793%

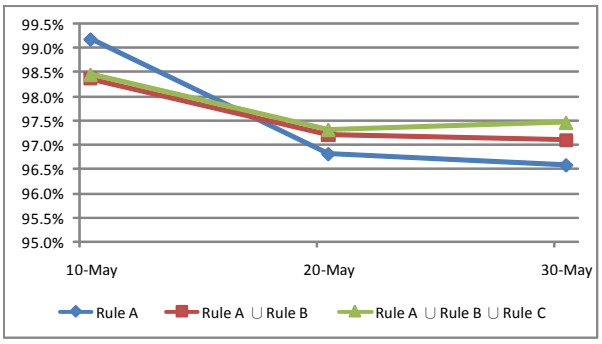
Experiment Result- Spam Recall

	10-May	20-May	30-May
Rule A	99.1640535%	96.8478261%	96.1423221%
Rule A ∪Rule B	99.3730408%	97.7173913%	98.4831461%
Rule A ∪Rule B ∪Rule C	99.4775340%	98.2608696%	99.2322097%

Experiment Result- Miss Rate

	10-May	20-May	30-May
Rule A	0.7656757%	12.7906977%	9.3333333%
Rule A ∪Rule B	8.1081081%	12.7906977%	8.0000000%
Rule A ∪Rule B ∪Rule C	8.1081081%	12.7906977%	6.6666667%

Experiment Result- Accuracy



	10-May	20-May	30-May
Rule A	99.1855204%	96.0238569%	96.4391951%
Rule A ∪Rule B	98.3710407%	96.8190855%	98.8713911%
Rule A ∪Rule B ∪Rule C	98.4615385%	97.3161034%	99.5013123%

Conclusion

- Due to rule exchange and evolution, collaborative approach is better than standalone server
- Collaborative approach can extend to hierarchical architecture
  - Some powerful server generate and exchange spam rules and spam rules can be transmitted to other powerless server
- In future study, spam rules can be generated by different rule-based approach and an integrated scheme will be developed

Q&A

Thank You!!