ATRC Workshop

TECHNICAL MEASURES TO COMBAT SPAM

Presented by:
James Seng
Assistant Director, Enabler Technologies
Infocomm Development Authority of Singapore

3 May 2005

What is spam?

- Definition of “spam” is controversial
  - Everyone defines it differently
  - “I know it when I see it”

- Common definition:
  - Bulk Unsolicited Commercial Email

General Strategy to stop spam

- Via economic & legal disincentives
  - Increasing the cost of sending spam
  - Reducing the effectiveness of spam

- Via Technical Means
  - Blacklisting, Whitelisting, Greylisting, etc.

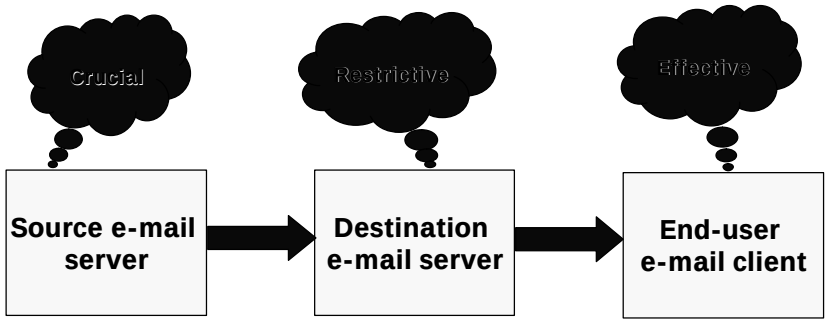
The reality

- **There is no silver bullet to curb spam**

- On their own, every measure has been circumvented by spammers:
  - Keyword filters ✍ Cleverly mis-spelled words
  - Adaptive filters ✍ Append legitimate tokens to confuse the filters
  - Blacklisting ✍ Domain spoofing / register with new ISPs
  - Legislation ✍ Legal loopholes / move operation to spam haven

- A multi-pronged approach comprising **legislation**, **technical measures**, **public education**, **industry self-regulation** & **international co-operation** is the "best way" of tackling the problem

Overview of technical measures



- Rate Limiting
- Authentication
- Payment
- Blocking Port 25

- Rate Limiting
- Reputation System
- Checksum

- Static Filtering
- Adaptive Filtering
- Reputation System

At source email server

- Most critical battlefield in the fight against spam
  - Win the war here, and we don't have to bear the cost of handling spam any further down the chain
- Unfortunately, it is a difficult war because
  - Too many e-mail servers
  - Some of them are operated by the spammers
  - Will take a long time for any one solution to be implemented by everyone
- Goal here is to get the "good guys" to make the change first
  - "Bad guys" who don't make the change risk getting blocked by everyone else

Possible measures at the source

Rate Limiting

- Limit number of outgoing e-mail (e.g. 100 e-mails per day per account)
- Already a fairly common practice

Authentication

- Prevent spammers from masquerading as you
- Authentication system by itself does not stop spam but complements other anti-spam measures

Payment

- Charge for sending e-mails
- Difficult to implement – need billing infrastructure, business peering etc.

Blocking Port 25

- Prevent "zombies" from sending out spam
- Can create problems for those who need to run their own e-mail servers

At the destination email server

- Important
  - Mailbox storage is limited – if mailbox is full, legitimate e-mail may get discarded
  - Bandwidth is limited – Users don't want to download too many e-mails to their client (especially mobile client)
- Yet our options are restricted
  - Tolerance level for spam higher than tolerance for loss of legitimate e-mails
  - One man's spam maybe another man's meat
- Hence we need to adopt a more conservative approach (e.g. tag/quarantine instead of discard spam)

Possible measures at the destination

Rate Limiting

- Limit number of incoming e-mail (e.g. 100 e-mails per day per account)

Reputation System

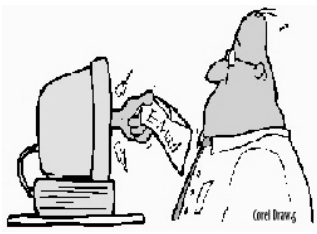
- Decide what to accept based on the known past reputation of the source e-mail server
- Whitelist based on business peering relationship
- Blacklist based on known spam servers, open relay servers
- Blacklist should be use with extreme care – may cause good e-mails to be discarded

Checksum

- Server generally can t tell if an e-mail is spam
- It can suspect an e-mail is spam if it is also sent to many other users on the server
- If similar e-mail is sent to many users in many servers, very high likelihood it is spam
- Compare checksum instead of actual content for privacy reasons

At the end user client

- Most flexible and effective
- High degree of control over their own e-mails
- Can achieve astonishing results with off-the-shelf software available today



Possible measures for the end user

Static Filtering

- Filter spam based on incoming email attributes, such as: sender name, subject title, email content etc.
- Need to constantly refine the filtering rules

Adaptive Filtering

- Filter spam based on statistical model
- Ability to learn what is spam and is non-spam based on user preference
- Most popular approach: Bayesian filtering can achieve up to 90% accuracy!!!

Reputation System

- Accept/reject emails based on incoming email address
- Blacklist generally not effective – each spam likely to come from different email address
- Whitelist – allowing only people in your address book to send your email

However…

- None of these measures are perfect
- But they all contribute to the battle against spam
  - Less than 10% of spam makes it to the end-user inbox
  - Blocking of port 25 and server-side authentication will significantly reduce the amount of spam that enters our networks
- All these measures (technical + legal + other approaches) do not need to add up to 100% effectiveness
- Because spammers will go away once the spam business is no longer commercially viable
- Let's be **optimistic!**

The Ultimate Solution

- There are always new technologies to deal with spam
  - "Blacklist is THE solution to spam" – 1998
  - "Challenge-response is THE solution to spam" – 2000
  - "Authentication is THE solution to spam" – 2002
  - "Reputation is THE solution to spam" – 2004
- The ultimate solution?

An email system that no one uses is also one that has no spam

Emerging Technologies

- Changes in communication technologies
  - Postal Mail, Telegraph, Ham Radio, Telephone, Fax, … Email, …
- Email is not the end of this technology evolution
  - Would we still use Email (as we know it) 10 years from now?

Communication & Spam

- "Spam" exist before Email and will continue to exist after Email
  - Junk Mail, Junk Fax, Direct Marketing, …
- It is fundamentally an economic game
  - Cost of marketing vs. return on investment
  - The lower the cost of communication, the better the RoI => more "spam".

Possible future "spam" problem

- Mobile Phone
  - SMS Spam
  - Mobile Email Spam
- Real-Time communication
  - Instant Messaging Spam
- VoIP (and lower Voice communication)
  - Voice Mail Spam
  - International Junk Fax
  - International Direct Marketing
- Blog/Wiki
  - Comment Spam
  - Wiki Spam

Only Email spam?

- Questions?
  - Should we focus *only* on “Email spam”?
  - Should we look beyond and develop a longer term solution?
- Well…
  - none of the “possible spam” has reach a critical stage like Email spam
  - Antispam solutions for those communication channel would be different Email
- So yes, lets focus only on Email for now.

What about long term solution?

There is no long term solution
(*in the long term, there is no Email*)

- But the experience and network establish today would be very valuable in the coordination of future spam.

Thank You