Update on the anti spam system at CERN

Pawel Grzywaczewski, CERN IT/OIS
HEPIX fall 2010

Current mail infrastructure

- Mail service in numbers:
  - ~18 000 mailboxes
  - ~ 18 000 mailing lists (e-groups)
  - ~ 3000 electronic faxes received/send each month
  - ~ 200 000 messages send each day to Internet
  - ~ 80 000 legitimate messages received each day from Internet
- Current systems
  - Microsoft Exchange 2003 – 95% of mailboxes
  - Microsoft Exchange 2010 – 5% of mailboxes + client access and edge servers

Overview

- Spam is a big problem, about 180 billion spam messages per day world wide
- At CERN each day we receive more than 1 million messages
- At CERN half of CERN mail addresses forward to external mail addresses
  - Large number of mailing lists
  - SPAM messages has to be rejected
  - Messages are **never** deleted – we refuse or accept
- New anti spam system reduced blacklisting of our mail servers

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

A bit of history about anti spam systems at CERN

2002 – home made anti spam system

Microsoft Exchange Server 2007

2008 – built-in system of Exchange 2007

Microsoft Forefront Protection 2010 for Exchange Server

April 2010 – Microsoft Forefront Protection 2010 for Exchange servers

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Motivation for a new anti spam system

- Users were reporting spam messages delivered to Inbox, Junk Folder
- Preparation for deployment of Exchange 2010 – new mail gateways
- System which is well integrated with current systems (possibility to whitelist from mailbox, whitelisting applied on all levels of filtering)

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

ForeFront and main features

- Anti spam system
- On board an efficient content filtering engine provided by Cloudmark
  - Fingerprints mechanism
    - Fingerprints of messages compared with fingerprints of known spam messages
    - Fingerprints calculated based on RELEVANT parts of a message
  - Heuristic
- Built in anti virus system
  - 5 different antivirus engines scan messages in parallel

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Sample spam statistics

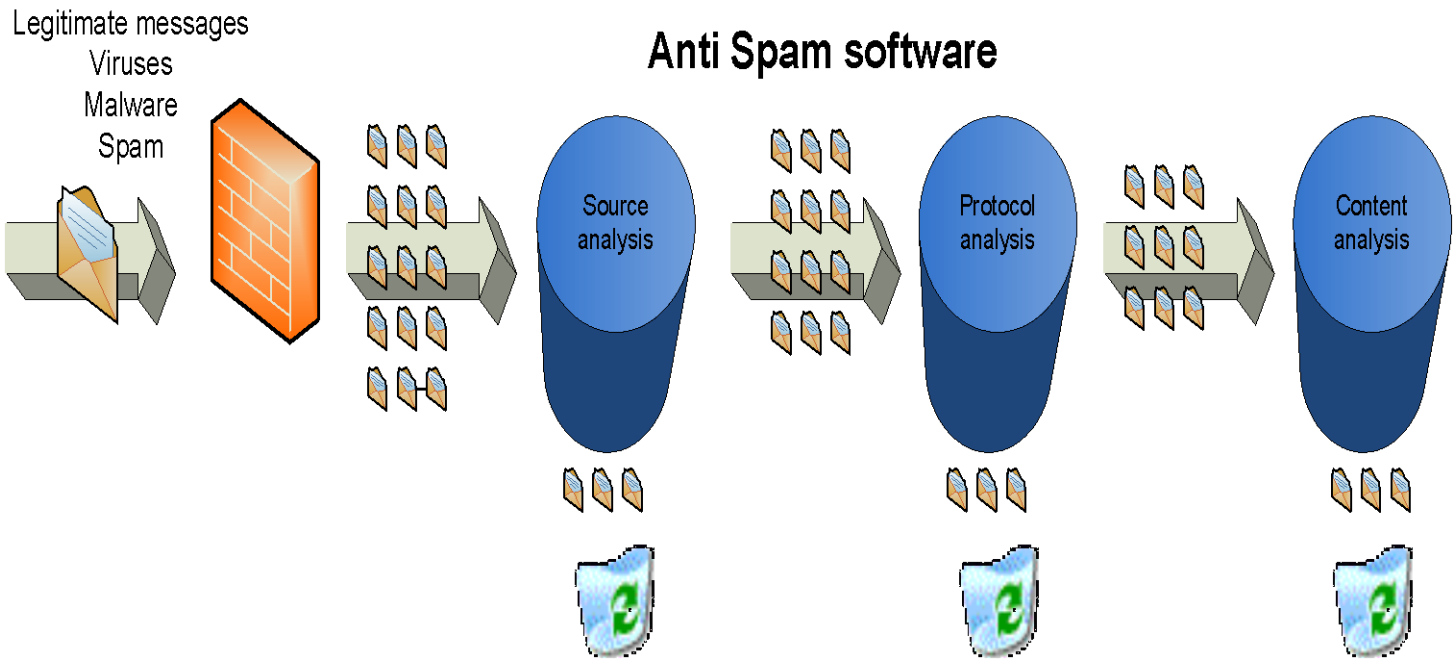
Mail Spam Statistics (Updated on 2010-10-16 13:42:42)		
Number of messages processed by anti spam filters (yesterday)		
Incomming mails	1219864	100%
Rejected	1149864	94.26%
Moved to spam folder	190	0.01%
Good Mail	69810	5.72%
Outgoing Mails	198755	

- 94% of messages is filtered
- Very low number of messages delivered to Junk Folder
- Helpdesk@cern.ch reported that amount of users complaining about spam decreased significantly
- Few requests from users who would like to receive more spam ;)

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Anti spam system architecture

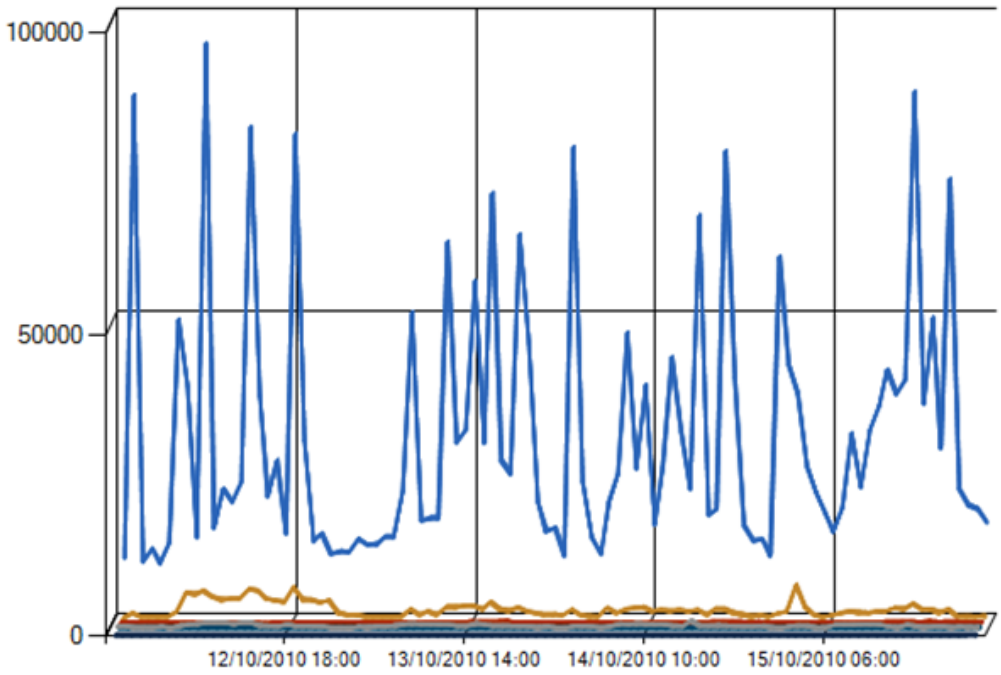
- 3 levels of spam filtering



- Most of messages is rejected by the first level

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Cause of messages rejection



Spam statistics - cause of message rejection

Name	Color
IP Block List Providers	
Recipient Filtering	
Sender Filtering	
Backscatter	
Content Filtering	
Black listed Internal Domain	

- 1st layer – source analysis: 1056475, ~94% of rejection
- 2nd layer – protocol analysis: 44198, ~4% of rejection
- 3rd layer – content analysis: 22455, ~2% of rejection

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

1st layer of defense – source analysis

Tell me what is your IP and I will tell you if you are SPAMMER

DNS block list

- One of the most effective means to counter spam attacks
- Forefront retrieves from public databases lists of blocked IP addresses
- Different providers: SPAMHAUS, Hotmail etc.

Sender id framework

- Prevent false positives

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

2nd layer - protocol analysis

Accept only legitimate recipients and senders

Sender filtering	Recipient filtering	Low percentage of false positives and false negatives	~4% of rejected messages
• Rejecting blocked senders • Accepting white listed	• Rejecting recipients which doesn't exist at CERN		

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Content analysis – 3rd layer

Why there are no messages in my Junk folder

Two methodologies are applied on messages • ***Fingerprint*** of a spam message (updates are received each 45 seconds) • ***Heuristic***	At present only few hundreds messages per day are delivered to spam folder, over all 18 000 mailboxes	2% of rejected messages

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Content filtering engine and fingerprints

Enterprise

Service providers

Mobile operator

Consumer

Cloudmark system

Nomination DB

Trust Evaluation System

Catalog Database

ForeFront with Cloudmark

Updates each 45 sec

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Remaining challenges

- Phishing
  - Looking for a solution
- Troubleshooting of messages rejected by Content Filtering
  - Improved by migrating users to Exchange 2010 – white listing
- Forwarding messages from mailboxes in other institutes
  - Improved by migrating users to Exchange 2010 – mailbox merge

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

What will come next

- When mailboxes are migrated to Exchange 2010
  - Possibility to whitelist senders
    - Whitelisting will be applied on all layers of filtering
  - By default all contacts will be whitelisted.
  - Blacklisting
  - Control of compromised accounts
    - Limits on number of recipients per 24h

Thank you!

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it