



# Spam Fighting at CERN



# What is Spam ?

- ◆ Spam is the friendly name given to unsolicited mail everyone receives in the mailbox.
- ◆ Comes from a Monty Python sketch, where in a café everything on the menu includes SPAM<sup>TM</sup> luncheon meat.
- ◆ Estimated cost for companies:
  - ◆ 1 spam = 1\$ cost per company (investment in spam fighting, helpdesk handling user complaints, time spent cleaning email folders...)
- ◆ Cost for spammers:
  - ◆ 39\$ for 1 million French email addresses.





# Email stealing

- ◆ **Test at CERN: an email address was published on the Mail Service Website, 37 days after the first Spam was received.**
- ◆ **6 Weeks study: 275 email addresses published on 175 different supports.**  
(source Federal Trade Commission, November 2002)
- ◆ **In 6 weeks: 3349 Spams were received by the 275 addresses.**
- ◆ **Speed record: First Spam was received 9 minutes after publishing an email in a Chat room.**

Support	Spammed emails
Chat room	100%
Newsgroup	86%
Standard Web site	86%
Personal Web Site	50%
Forum	27%
WebMail	9%

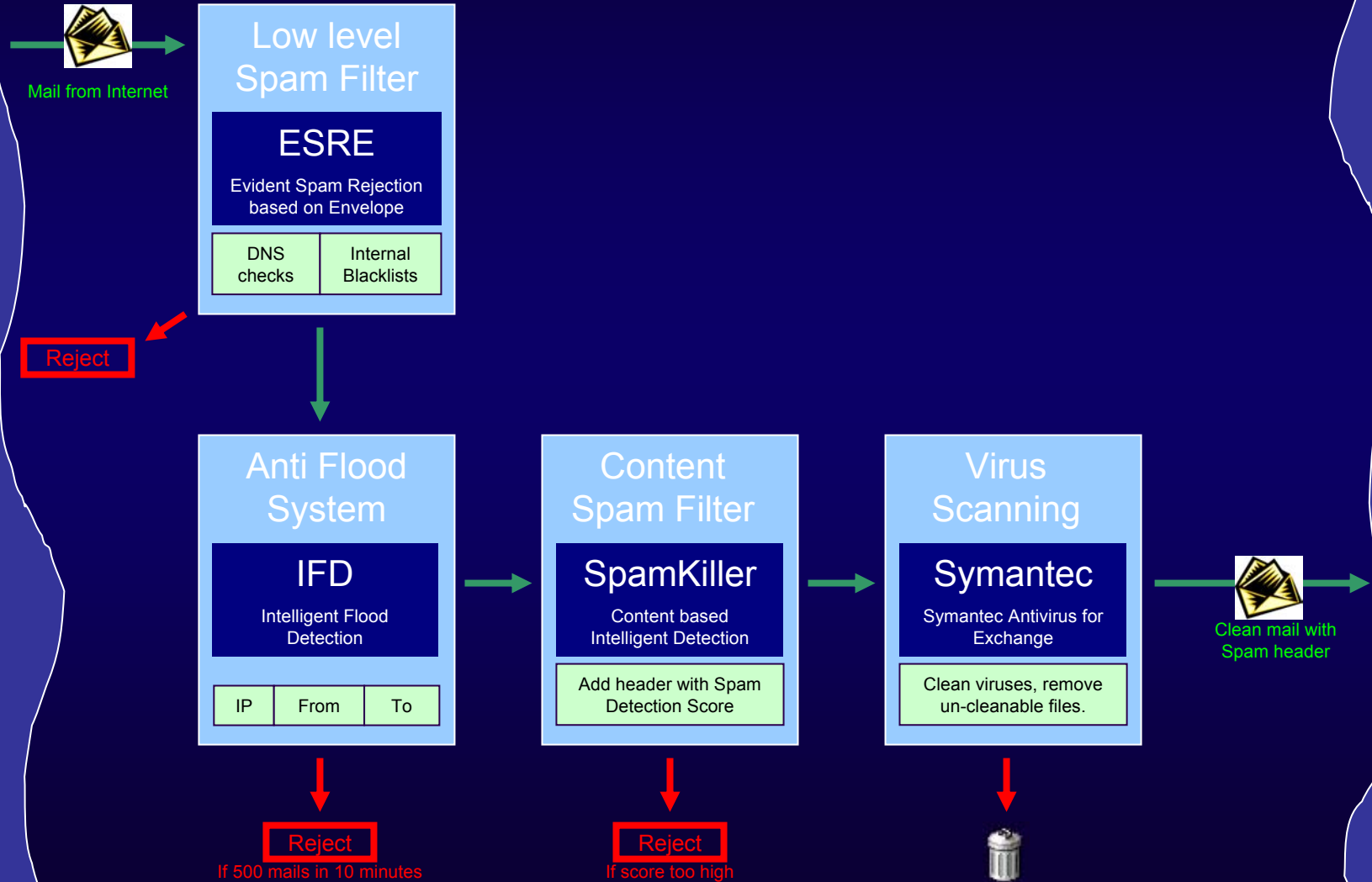


# Products review

- ◆ Existing market products were reviewed:
  - ◆ Technology too young
  - ◆ Results are not accurate
  - ◆ Missing a per user basis configuration
- ◆ While the market consolidates ...
  - ◆ CERN/IT developed its own Anti-Spam filter.
  - ◆ Less effort than running after immature commercial technology.
  - ◆ Now running for 1.5 year.
  - ◆ Easy to modify and update detection techniques.
  - ◆ CERN specific user level configuration / customization.

# Mail filtering overview

Internet / Outside CERN



Exchange Back-Ends / Other CERN Mail Servers



# Content Spam Filtering

- ◆ **CERN SpamKiller is NOT McAfee Spamkiller.**
- ◆ **SpamKiller calculates the probability for a message to be spam**
  - ◆ **Regular expressions.**
  - ◆ **“Intelligent” content parsing.**
  - ◆ **Statistical heuristics (Bayesian Filters).**
  - ◆ **Charset detection algorithm.**
- ◆ **The user sets the threshold at which he wants spam to be rejected**
  - ◆ **Rejected message can be seen by the user (CERN Spam folder)**
  - ◆ **Per user configuration**
  - ◆ **Rejection of foreign languages mail on a per user basis (Chinese, Korean, Russian, Japanese, Arabic, etc ...)**



# User configuration

## Spam Fight

The MMM Spam Filter analyses incoming mail and moves those identified as Spam to the folder Cern Spam.

Please check this folder occasionally for mail that has been classified as Spam.

**Filtering level**

### Select the filtering level:

- Off** No Spam filtering, all incoming mails will be delivered without filtering.
- Low** Obvious Spam mail will be detected and moved to the Cern Spam folder, but some will still arrive in your Inbox.
- Medium** Obvious and more "intelligent" Spam mail will be detected and moved to the Cern Spam folder.
- High** Nearly all Spam mail will be detected and moved to the Cern Spam folder. Some commercial mailings might be assimilated as Spam, you'll need to check the Cern Spam folder to identify and whitelist them.  
*Recommended*

**Expiration** Keep spam filtered mails for :

Delete automatically Evident Spam without moving them to Cern Spam folder.

Switch to: [\[Simple mode\]](#)[\[Advanced Mode\]](#)

## Whitelist

Defines trusted mail sources to bypass the Spam filter. You can add individuals, sites or mailing lists based on the sender, recipient or subject content.

### Header String to match

<a href="#">Edit</a> <a href="#">Delete</a> From	superfourmi
<a href="#">Edit</a> <a href="#">Delete</a> From	@240z.org
<a href="#">Edit</a> <a href="#">Delete</a> From	@ldlc.com
<a href="#">Edit</a> <a href="#">Delete</a> From	@dvdrama.com
<a href="#">Edit</a> <a href="#">Delete</a> From	@google.com

Add all MMM Contacts to whitelist: [\[Add Contacts\]](#)

## Foreign languages based on character sets

If you receive mail in foreign character sets (i.e. Chinese), you may want them to be treated as Spam.

Note that for example Italian or German languages are part of the "western" character set and can therefore not be filtered.

### Move to Cern Spam

- Chinese mails
- Korean mails
- Japanese mails
- Russian mails

[\[Hide details\]](#)[\[Show details\]](#)

**Language-based rejection**



# Efficiency

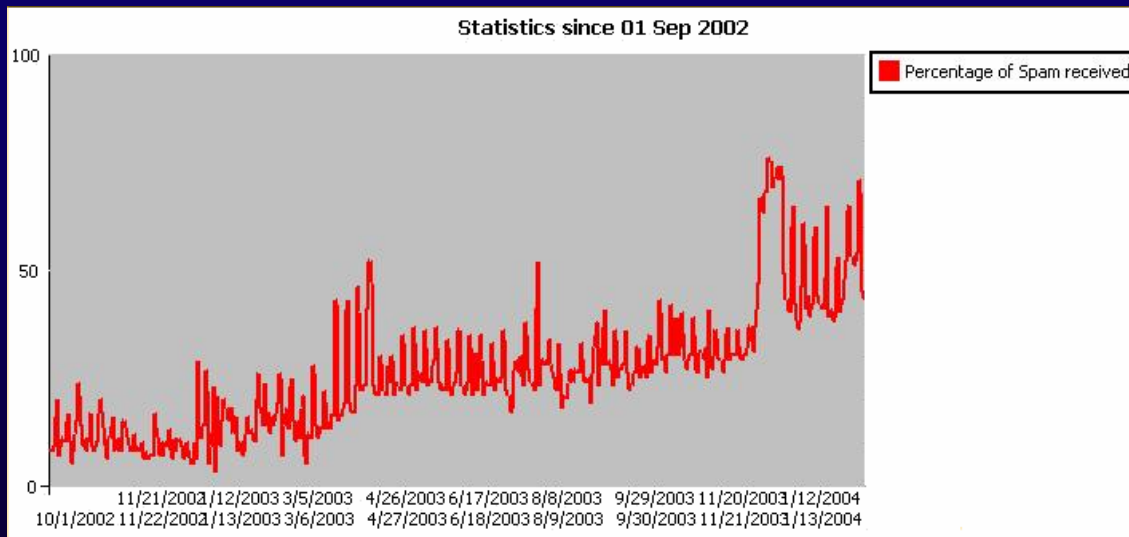
Smtip Traffic (yesterday): [\[ Details \]](#)

Reject	In	Out
120259	98820	14259
Spam: 59020		
Spam: 179279 (81%)		

1 day statistics on smtp gateways, all checks enabled:

**CERN receives 81% of Spam ! But 67% is rejected.**

More than 50% of accepted traffic is detected as spam.







# Efficiency

- ◆ **False positives are quite low**
  - ◆ **Except for commercial lists (spam that you want).**
  - ◆ **White lists at user level can be configured to prevent this.**
- ◆ **Good spam detection**
  - ◆ **My mailbox filtering is standard:**
    - ◆ 30 to 40 Spams filtered per day.
    - ◆ 3 or 4 Spams still go to the INBOX per week.
  - ◆ **Can be improved, but new algorithms must be found.**
- ◆ **Not enough for some users with “public” email address**
  - ◆ **Old email address or published email address are more targeted for Spam.**

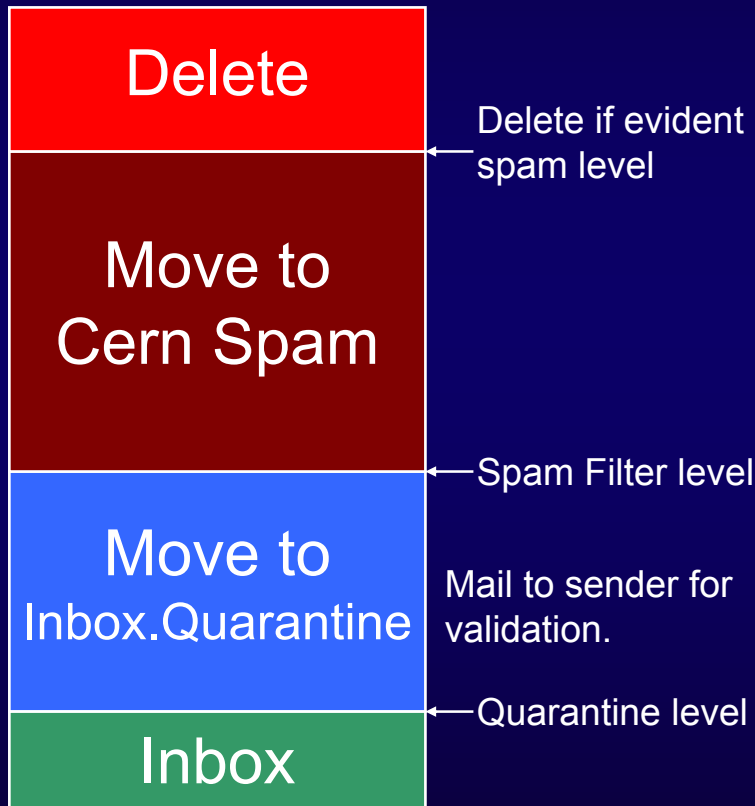


# Future evolution

- ◆ **Spammer techniques always follow anti-spam techniques.**
- ◆ **New detection mechanisms work only for a few months.**
- ◆ **Needs a full time work to have a constantly “up-to-date” filter.**
- ◆ **Only viable long term solution is to accept only mails from people you know:**
  - ◆ **ICQ (and other messenger systems) already have this feature.**
  - ◆ **Accept only messages from people in my contact list.**
  - ◆ **Adding someone to the contact list requires validation.**



# New feature (in test)



- ◆ Good Mails not matching the user's whitelist are quarantined.
- ◆ Mail is send to sender requiring action to validate himself.

You recently sent an mail to "ivan4@cernxchg.cern.ch" .  
However, the personal CERN Spam Filter of this user doesn't recognize you and needs to verify your email address to complete delivery of the mail.

To complete the delivery of your mail, please click on the link below:  
[Click here to complete delivery of the mail](#)

#### Email address validation

To validate email address emmanuel.ormancey@cern.ch copy the code below in the field and press button:

4587

UnBlock

- ◆ Once validated, sender is added to whitelist, mails are moved back to Inbox.



# Next...

- ◆ **Current situation:**
  - ◆ Think, test and add new techniques.
  - ◆ Improve a fully customizable solution at user level.
- ◆ **Improvements**
  - ◆ Automatic whitelist currently in test.
- ◆ **Future is to join forces against Spam:**
  - ◆ Share rules, regular expressions patterns and Bayesian statistics dictionary with other organizations.
  - ◆ Central Antispam configuration with Live Update like antivirus definitions will be the solution. Therefore ...
- ◆ **Long term goal: use a commercial product.**
  - ◆ Like for antivirus products, only a full time working team will provide up-to-date filters.



# Questions ?

[emmanuel.ormancey@cern.ch](mailto:emmanuel.ormancey@cern.ch)