Spam Filtering at CERN

Emmanuel Ormancey - 23 October 2002

Topics

- Statistics
- Current Spam filtering at CERN
- Products overview
- Selected solution
- How it works
- Exchange 2000 integration

Some statistics…

- At CERN:
  - Low level existing filters: 25% of mails detected as spam and rejected.
  - New filtering solution identifies 10% more.
- Measurements in Europe for 2001 (NetValue users panel) :
  - Spam increased of 80% in 2001.
  - 36.8% of received mails are Spam.
- According to US AntiSpam company Brightmail:
  - Spam increased of 450% during last year
  - 74% of received mails are Spam.

Current Spam Filtering

- Basic checks:
  - Sendmail level tests.
  - Local lists of banned IP addresses, domains, subject keywords, emails.
  - Header “consistency” tests (i.e. message id format).
- Mail rejected if identified as Spam.
- Manual work:
  - Update local banned lists from abuse reports.
  - Remove entries when users report false positive rejections.

Commercial products

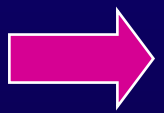
- **Commercial products too basic**
  - **Basic tests:**
    - keywords in subject/body
    - IP address ban
    - Sender / recipient ban
  - **Action:**
    - Delete: helpdesk will receive user complaints if false positive.
    - Quarantine (i.e. Norton antivirus): require manual lookup to validate real spam and good mails.

SpamAssassin testing

- How it works:
  - All in one: Different tests based on different techniques
  - Client / server version, with a ‘simple client’ allowing portability.
- Good for spam detection.
- Stability problem (on our Solaris).
- Need to correct regular expressions bugs.
- Not enough, need a mix of:
  - Mail content tests (SpamAssassin)
  - Low level “sendmail” tests (actual spam tests)
- Need some custom rules and tests.
- Need logs and statistics.

Solution

- Start from SpamAssassin base
- Add existing rules and custom tests
- Easy to modify and to create add-ins.
- Windows based: Future Exchange 2000



C# .NET SpamKiller

- Easy to develop in any language.
- Compiled regular expressions, compatible with unix.
- After 3 months running and stress testing: no crash, no leak: seems stable.

Detecting spam - Tests

- **Different tests:**
  - **Text only (regular expressions):**
    - Header
    - Body full text
    - Body raw for base64 encoded spam
  - **“Smart tests” more complex than regular expressions.**
  - **Header consistency.**
  - **Open relays blacklist check on several servers.**
  - **Catalog check: compares mail with spam catalog (calculated signatures and subjects keywords).**

Detecting spam – Scoring

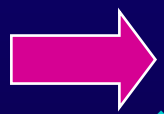
- **Score calculation:**
  - **Each test returning true returns a score.**
  - **If sum of all scores is greater than 'required hits', mail is spam.**
  - **Lowest 'required hits' value is 5.**

Sample:

Spam: True ; 5.559 / 5
Content analysis details: (5.559 hits, 5 required)
2 points: HTML-only mail, with no text version
0.21 points: 'Received:' has 'may be forged' warning
0.814 points: Subject has an exclamation mark
0.5 points: Spam phrases score is 00 to 01 (low)
2.035 points: 'remove' URL contains an email address

Detecting spam - Action

- When spam is detected:
  - Do not delete mail, it may be an error or a commercial mailing list subscribed by user.
  - Do not reply to sender “we don’t accept spam” → it helps to improve spammer techniques.
  - Do not quarantine mail at server level: too much traffic and too much work.
  - A good mail service don’t loose mails.



- Solution: Let the user decide
  - Quarantine spam mail at the user level.
  - Allow user to check in quarantined mails for missing mails.
  - Allow user to choose a spam detection level (lowest level = 5)
  - Allow user to choose quarantine behavior.

User choice

Spam Fight

Cern Exchange 2000 Spam protection:

The Cern Spam Filter analyses incoming mails and move identified Spam mails to the folder **Cern Spam**.
Have a look to this folder **Cern Spam** from time to time to verify that no real mail has been moved.
To disable Spam filtering you can use this form or simply remove **Cern Spam** folder.

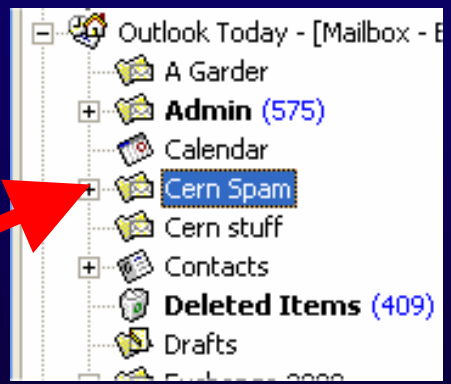
Different levels of filtering can be set:

○ **Off**	No Spam filtering, all incoming mails will be delivered without filtering.
○ **Low**	Evident Spam mails will be detected and moved to the Cern Spam folder, but some will still arrive in your inbox. The risk of identifying a "true mail" as spam is very low.
○ **Medium**	Evident and more "intelligent" Spam mails will be detected and moved to the Cern Spam folder. The risk of identifying a "true mail" as spam is low, except for some commercial mailing lists which are often bad formatted.
◉ **High**	Nearly ALL Spam mails will be detected and moved to the Cern Spam folder. The risk of identifying a "true mail" as spam is important, commercial mailing will often be assimilated as spam. You'll need to check the Cern Spam folder more often.
Expiration	Keep spam filtered mails for : Always

Save

- Configure Spam Level.
- Set expiration time.

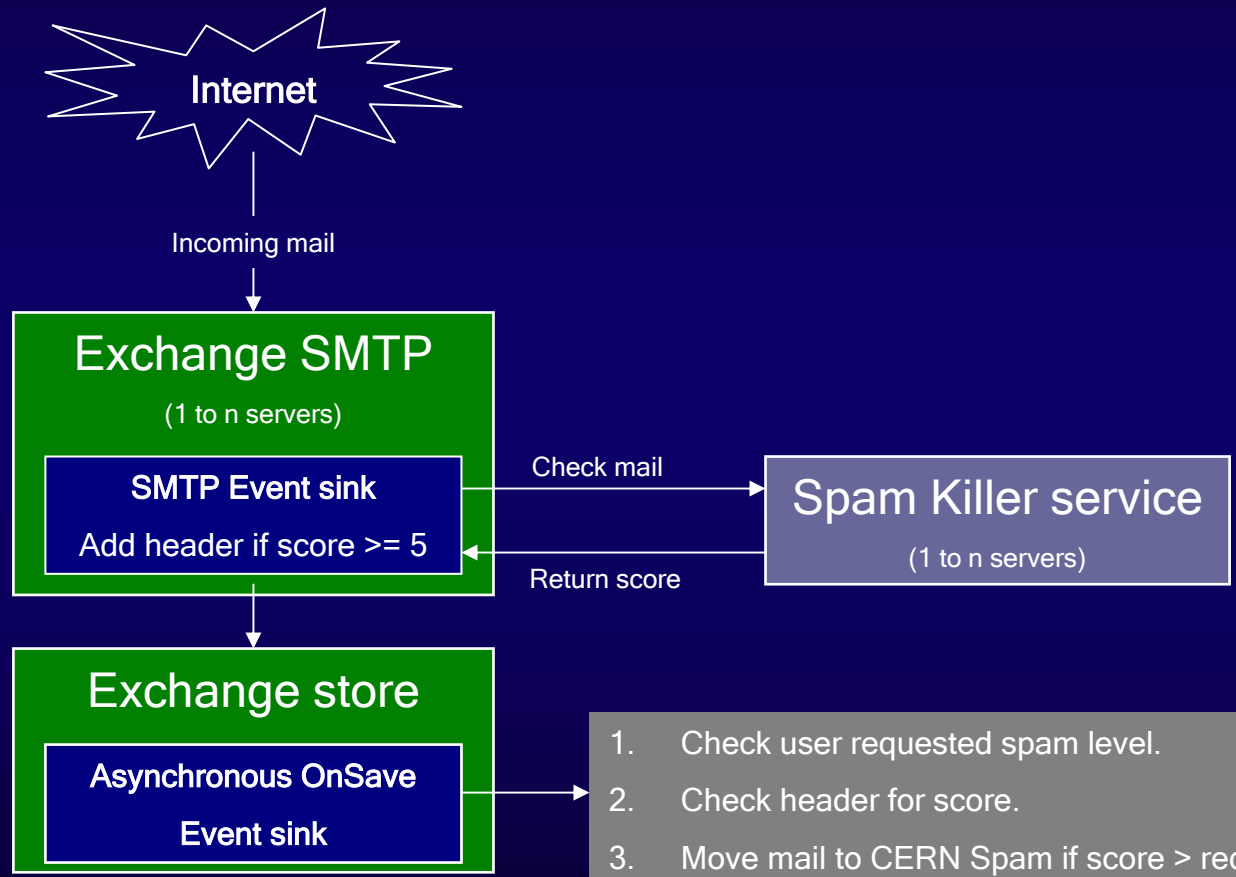
Cern Spam folder automatically created.



SpamKiller – Overview

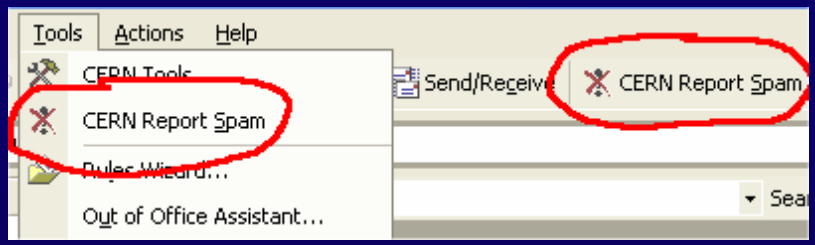
- Server:
  - Windows service.
  - Multithread "http like" server (clients on any platform can use it).
  - High exception catch to prevent server crash on error or bug.
- Configuration:
  - Configuration in XML files (import from original SpamAssassin configuration possible).
  - Precompiled regular expressions to gain performance.
- Statistics and logging:
  - Logs to perfmon (performance monitor) real-time statistics.
  - Logs statistics into XML files.

Exchange integration



Reporting Spam

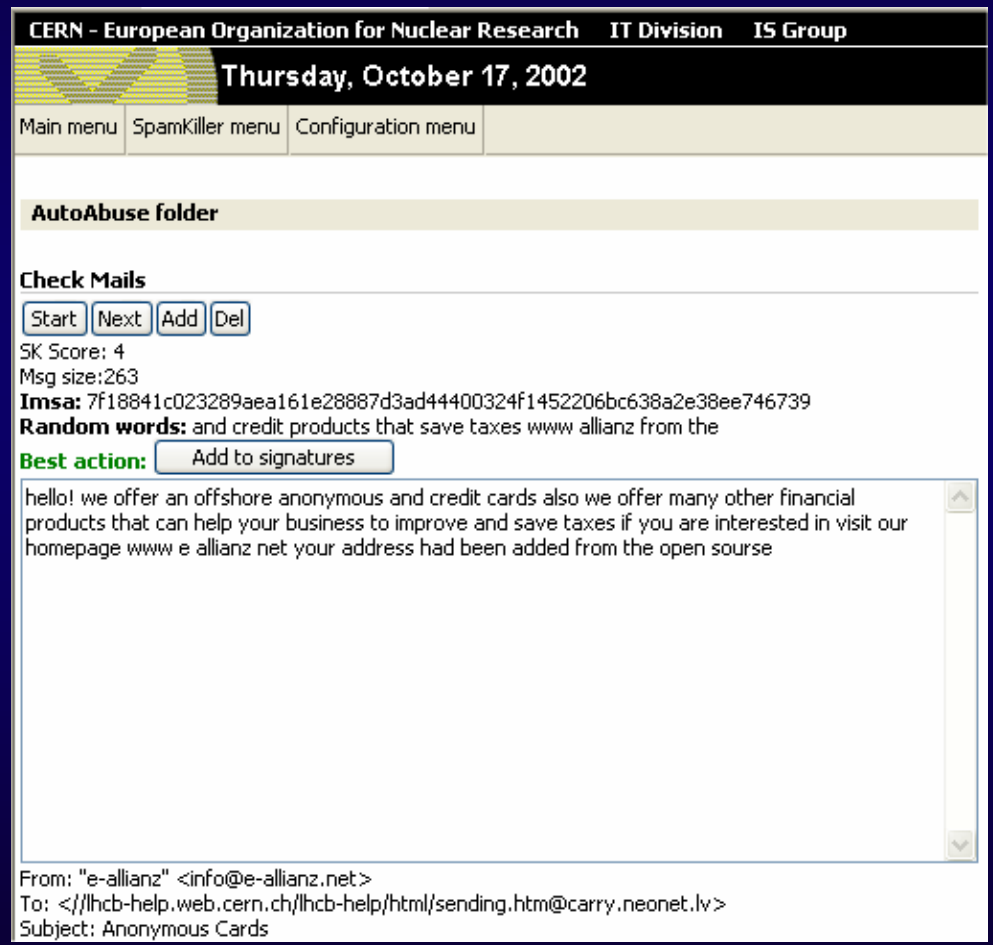
- Outlook XP: Com Add-in adds button to report spam (moves selected mails to specific public folder).



- Others: Forward mail to abuse@cern.ch

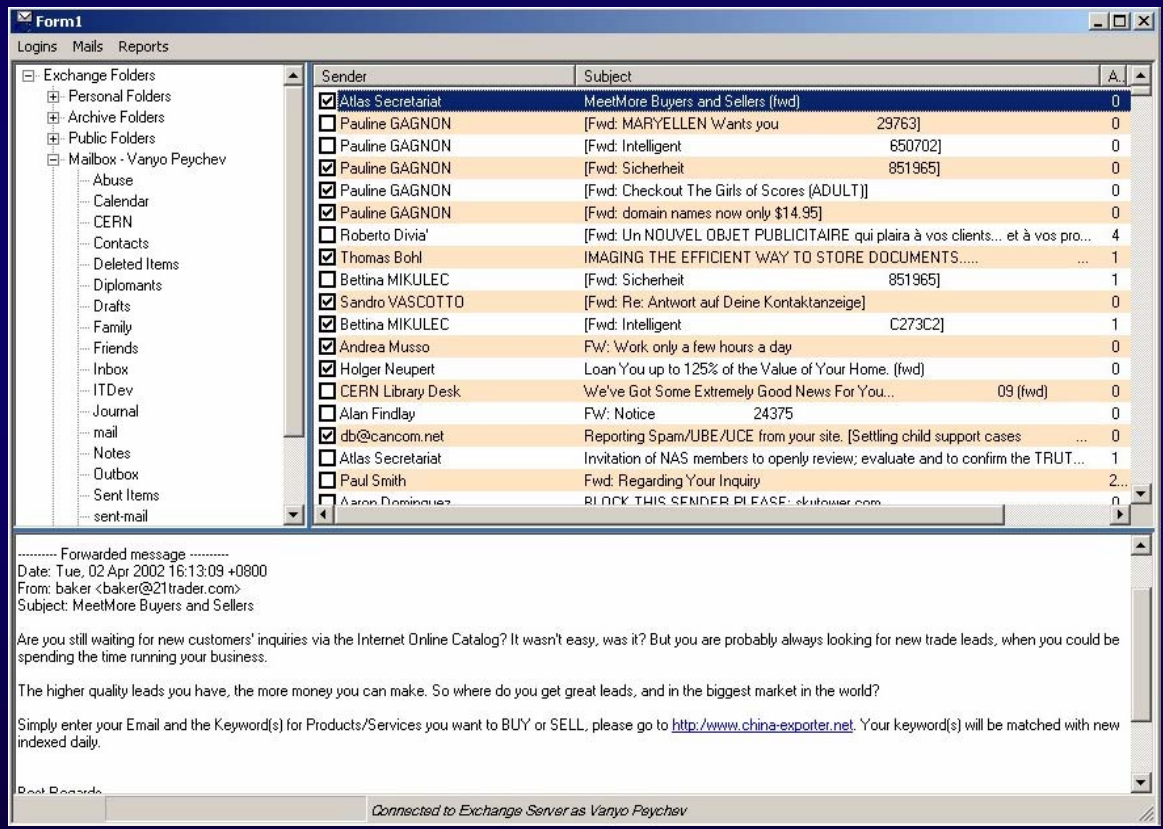
Use of reported Spam

- Spam reported with add-in button:
  - Mail in original format.
  - Create signatures.
  - Add signatures to catalog.
  - Can be automated.



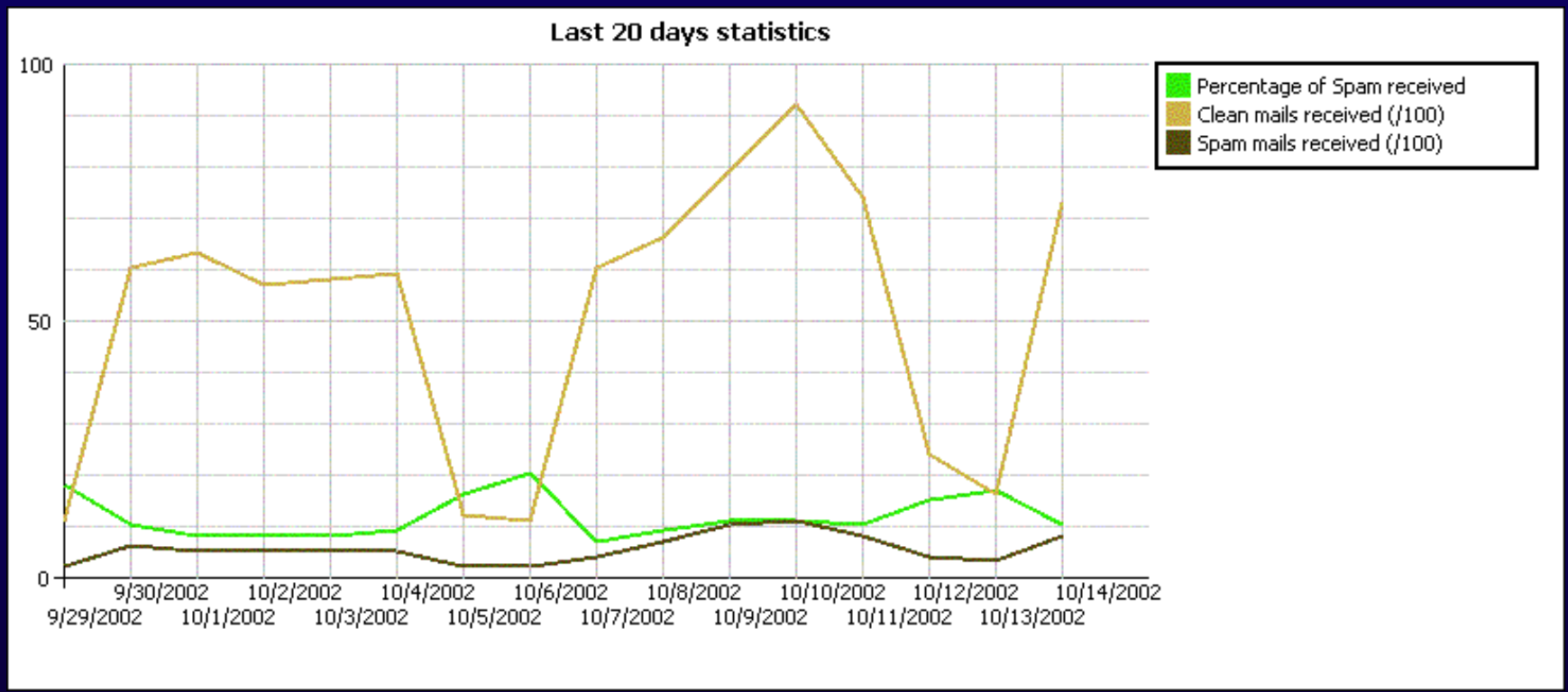
Use of reported Spam

- Spam forwarded to abuse@cern.ch
  - Mail modified due to forward.
  - Extract header information.
  - Create catalog:
    - Subjects
    - IP
    - Senders



Statistics

Online statistics available on SpamKiller website:



Conclusion

- Now available to CERN Exchange users.
- Up since July.
- Low manual work: populate Spam catalog with tools, tune rules.
- Problem with mailing lists filtering: add white list at user level in next release.
- Clients can be created on any system. (possible reuse of SpamAssassin client).

Questions ?

Contact: emmanuel.ormancey@cern.ch