



# 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.

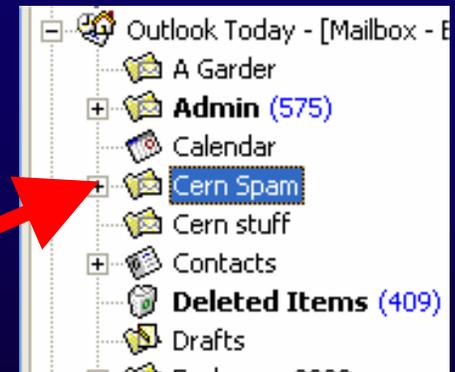
Different levels of filtering can be set:

|  |   |
|--|---|
| <input type="radio"/> <b>Off</b>             | No Spam filtering, all incoming mails will be delivered without filtering.  |
| <input type="radio"/> <b>Low</b>             | Evident Spam mails will be detected and moved to the <b>Cern Spam</b> folder, but some will still arrive in your inbox. The risk of identifying a "true mail" as spam is very low.  |
| <input type="radio"/> <b>Medium</b>          | Evident and more "intelligent" Spam mails will be detected and moved to the <b>Cern Spam</b> folder. The risk of identifying a "true mail" as spam is low, except for some commercial mailing lists which are often bad formatted.                              |
| <input checked="" type="radio"/> <b>High</b> | Nearly ALL Spam mails will be detected and moved to the <b>Cern Spam</b> 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 <b>Cern Spam</b> folder more often. |

**Expiration** Keep spam filtered mails for :  

- Configure Spam Level.
- Set expiration time.

Cern Spam folder automatically created.



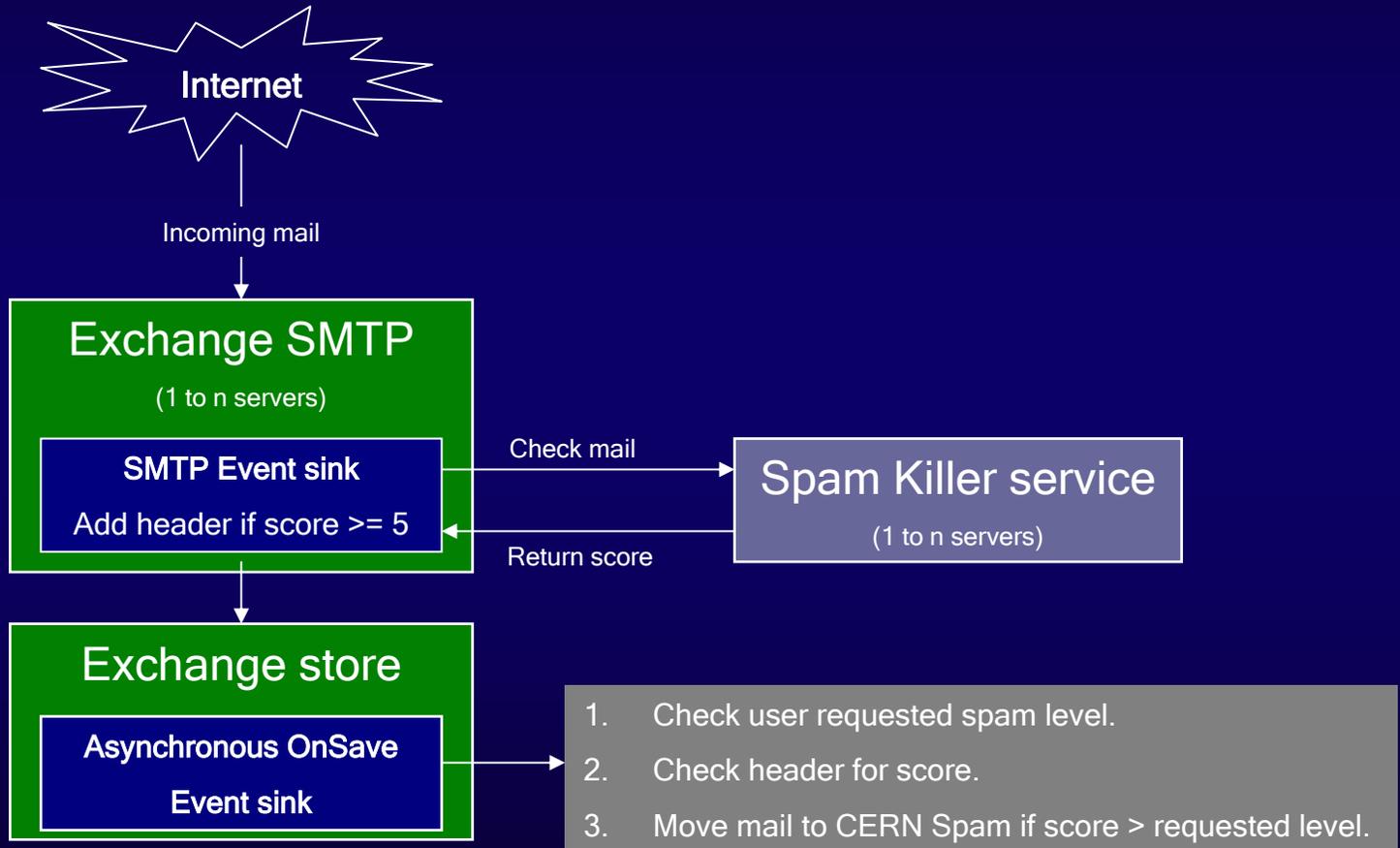


# SpamKiller – Overview

- ◆ **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.**



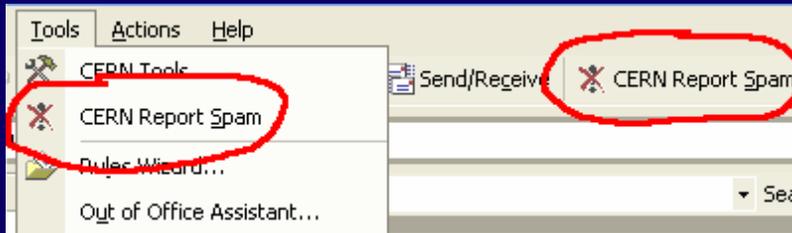
# Exchange integration





# 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](mailto: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.

CERN - European Organization for Nuclear Research IT Division IS Group

Thursday, October 17, 2002

Main menu SpamKiller menu Configuration menu

**AutoAbuse folder**

**Check Mails**

SK Score: 4  
Msg size:263  
**Imsa:** 7f18841c023289aea161e28887d3ad44400324f1452206bc638a2e38ee746739  
**Random words:** and credit products that save taxes www.allianz from the  
**Best action:**

hello! we offer an offshore anonymous and credit cards also we offer many other financial products that can help your business to improve and save taxes if you are interested in visit our homepage www.e.allianz.net your address had been added from the open source

From: "e-allianz" <info@e-allianz.net>  
To: <://lhcb-help.web.cern.ch/lhcb-help/html/sending.htm@carry.neonet.lv>  
Subject: Anonymous Cards



# Use of reported Spam

## ◆ Spam forwarded to [abuse@cern.ch](mailto:abuse@cern.ch)

- ◆ Mail modified due to forward.
- ◆ Extract header information.
- ◆ Create catalog:
  - ◆ Subjects
  - ◆ IP
  - ◆ Senders

The screenshot shows an Outlook window titled 'Form1' with a list of emails in the main pane. The left pane shows the folder structure, including 'Exchange Folders', 'Personal Folders', 'Archive Folders', 'Public Folders', and 'Mailbox - Vanyo Peychev'. The main pane displays a list of emails with columns for 'Sender' and 'Subject'. The preview pane shows a forwarded message from 'baker <baker@21trader.com>' with the subject 'MeetMore Buyers and Sellers'. The message content includes a promotional text about an Internet Online Catalog and a link to <http://www.china-exporter.net>.

| Sender  | Subject   | Count |
|---|---|-------|
| <input checked="" type="checkbox"/> Atlas Secretariat | MeetMore Buyers and Sellers (fwd)   | 0     |
| <input type="checkbox"/> Pauline GAGNON               | [Fwd: MARYELLEN Wants you 29763]  | 0     |
| <input type="checkbox"/> Pauline GAGNON               | [Fwd: Intelligent 650702]   | 0     |
| <input checked="" type="checkbox"/> Pauline GAGNON    | [Fwd: Sicherheit 851965]  | 0     |
| <input checked="" type="checkbox"/> Pauline GAGNON    | [Fwd: Checkout The Girls of Scores (ADULT)]                                     | 0     |
| <input checked="" type="checkbox"/> Pauline GAGNON    | [Fwd: domain names now only \$14.95]  | 0     |
| <input type="checkbox"/> Roberto Divia'               | [Fwd: Un NOUVEL OBJET PUBLICITAIRE qui plaira à vos clients... et à vos pro...  | 4     |
| <input checked="" type="checkbox"/> Thomas Bohl       | IMAGING THE EFFICIENT WAY TO STORE DOCUMENTS.....                               | 1     |
| <input type="checkbox"/> Bettina MIKULEC              | [Fwd: Sicherheit 851965]  | 1     |
| <input checked="" type="checkbox"/> Sandro VASCOTTO   | [Fwd: Re: Antwort auf Deine Kontaktanzeige]                                     | 0     |
| <input checked="" type="checkbox"/> Bettina MIKULEC   | [Fwd: Intelligent C273C2]   | 1     |
| <input checked="" type="checkbox"/> Andrea Musso      | FW: Work only a few hours a day   | 0     |
| <input checked="" type="checkbox"/> Holger Neupert    | Loan You up to 125% of the Value of Your Home. (fwd)                            | 0     |
| <input type="checkbox"/> CERN Library Desk            | We've Got Some Extremely Good News For You... 09 (fwd)                          | 0     |
| <input type="checkbox"/> Alan Findlay                 | FW: Notice 24375  | 0     |
| <input checked="" type="checkbox"/> db@cancom.net     | Reporting Spam/UBE/UCE from your site. [Settling child support cases ...        | 0     |
| <input type="checkbox"/> Atlas Secretariat            | Invitation of NAS members to openly review; evaluate and to confirm the TRUT... | 1     |
| <input type="checkbox"/> Paul Smith                   | Fwd: Regarding Your Inquiry   | 2...  |
| <input type="checkbox"/> Aaron Dominguez              | BLOCK THIS SENDER PLEASE: sktowner.com  | 0     |

----- Forwarded message -----  
Date: Tue, 02 Apr 2002 16:13:09 +0800  
From: baker <baker@21trader.com>  
Subject: MeetMore Buyers and Sellers

Are you still waiting for new customers' inquiries via the Internet Online Catalog? It wasn't easy, was it? But you are probably always looking for new trade leads, when you could be spending the time running your business.

The higher quality leads you have, the more money you can make. So where do you get great leads, and in the biggest market in the world?

Simply enter your Email and the Keyword(s) for Products/Services you want to BUY or SELL, please go to <http://www.china-exporter.net>. Your keyword(s) will be matched with new indexed daily.

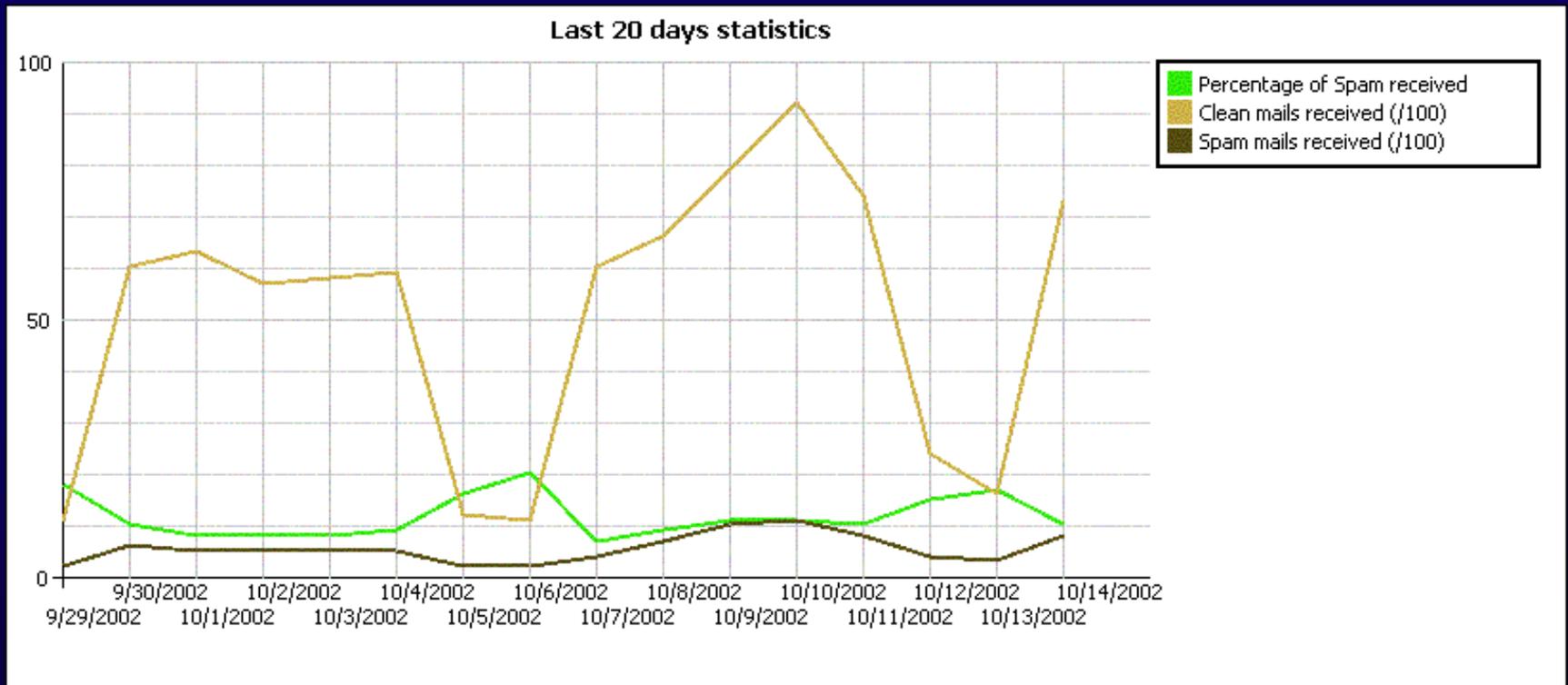
Best Regards,

Connected to Exchange Server as Vanyo Peychev



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