Emails: Spam and Viruses – how do we react

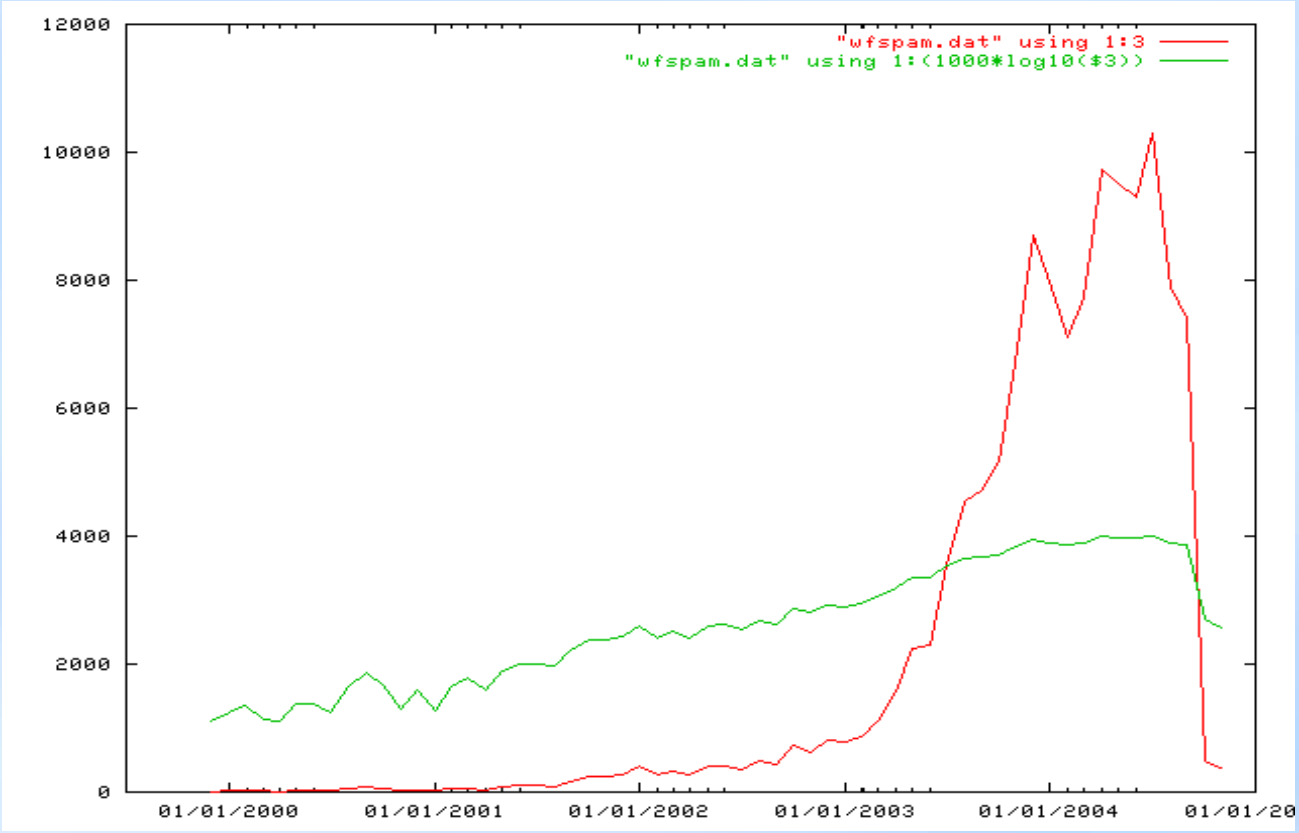
Technical Seminar

Wolfgang Friebel

Fight against spam is a hard job

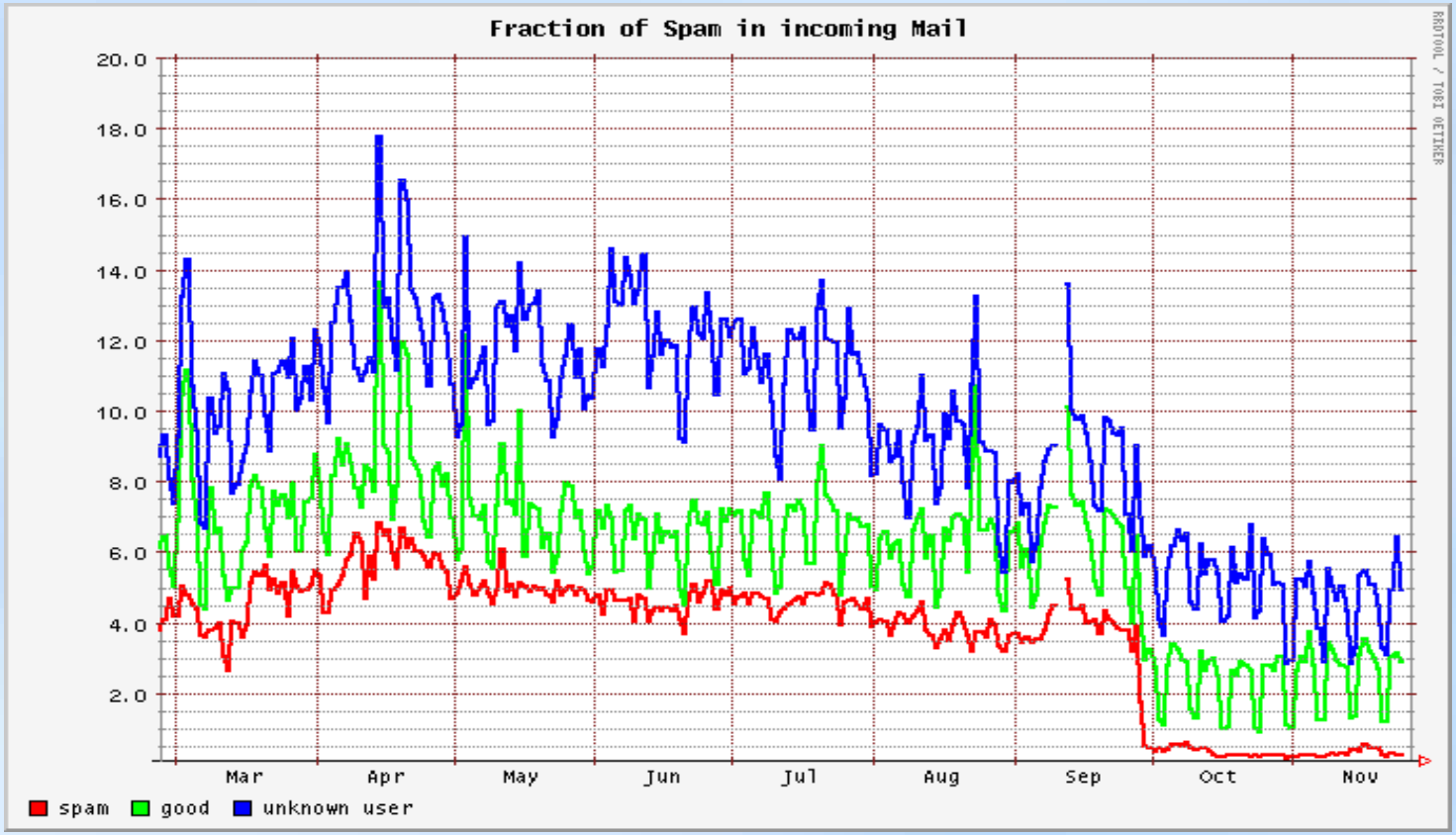
Spam and Viruses in Emails – current status

- monthly spam statistics of a specific account since the year 2000 (linear and log scale)



Fraction of Spam in incoming mail

- Majority of mails is again good mail since Sept, 27
  - Blue: old and nonexisting addresses being probed



Who is profiting from Spam

- Sending Spam is profitable
  - If you do not get sentenced

Spammer Jeremy Jaynes, who operated using the alias 'Gaven Stubberfield' and was listed by Spamhaus as the 8th most prolific spammer in the world, has been convicted of spamming using deceptive routing information to hide the source. A Virginia court recommended Jaynes spend nine years in prison for sending hundreds of thousands of unsolicited bulk emails. (Quelle: www.spamhaus.org)

- According to Forrester Research email-marketing is a 4.8 Billion Dollar business in 2003
- According to Ferris Research profit expectations in 2003 were
  - 130 Mio $ for companies selling anti-spam products
  - 20-30 Mio $ for Spammers

The worlds largest Spammers

- Source: www.spamhaus.org

Top 10 Spam Countries October 2004		Top 10 Worst Spam ISPs October 2004		Top 10 ROKSO Spammers October 2004	
1	United States	1	mci.com	1	Alan Ralsky
2	China	2	kornet.net	2	Yambo Financials
3	South Korea	3	chinanet-cq	3	Scott Richter - Wholesalebandwidth
4	Taiwan	4	above.net	4	Michael Lindsay / iMedia Networks
5	Canada	5	chinanet-gd	5	Alexey Panov - ckync.com
6	Brazil	6	comcast.net	6	Bill Waggoner
7	Russia	7	chinanet-sh	7	Robert Soloway - NIM
8	Japan	8	pacbell.net	8	Elmar Brunenieks
9	Argentina	9	xo.com	9	Eddy Marin - Oneroute
10	United Kingdom	10	verizon.net	10	Ibragimov Ruslan / send-safe.com
	Source: SBL database		Source: SBL database		Source: SBL + ROKSO database

Spam and Viruses

- Spam sender and virus ”producer” do often cooperate
- Viruses do prepare the infrastructure for spam attacks (SMTP Gateway)
- Article about the author of the Sobig virus
  - http://matthias.leisi.net/archives/92-guid.html
  - Most probable author of the virus is Ibragimov
  - See list of largest spammers
- Spam and virus spreading can be seen sometimes as a simultaneous attack of the mail servers from many sites (DDOS – distributed denial of service)

Typical history of a spam mail

- collection and selling of addresses
- Order to advertise a product by email
- Order will be carried out by a spammer
- Mass mailer sends emails directly or via relay
- Relay (hacked computer) redirects mail
- Mail hits DESY router
- Mail is getting to the mail server
- Mail is stored in the user's INBOX
- Mail is being read by a mail reader

Measures against spam

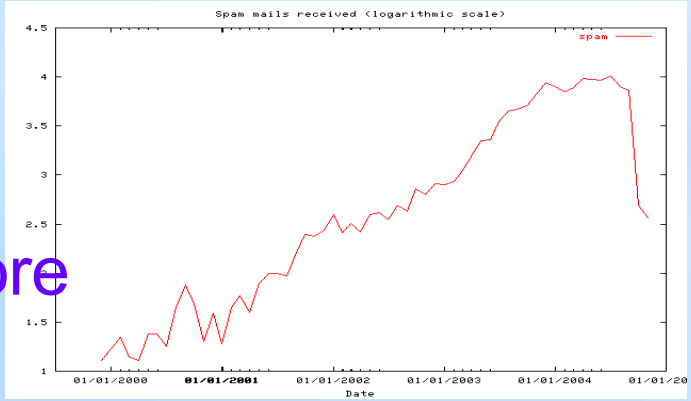
- At each point on the way of an email from the sender to the receiver action can be taken
- Ultimate goal: **100 percent of “good” mails will get delivered properly**
- All actions have to comply to that goal
- No spam fighting at any price
- Most efficient method:
  - Finding the causes for spam and fighting against it
  - Soothing of the symptoms is often much more costly

“Harvesting” of email addresses

- Make it more difficult
  - No publishing of address lists (LDAP for HEP only)
  - In Mails: signature w.f AT desy.de, **not** w.f@desy.de
  - In web pages: use a short Javascript program to generate mailto: URL (available from webmaster)
  - Do not react on dubious mails
  - Do not buy anything being advertised in spam mails!!
  - Do not follow unsubscribe requests in spam
- Do not use software of dubious origin
  - Spyware and software containing data mining code especially for Windows

Legal actions against senders

- Anti spam laws
  - Laws against Spam do exist only in few countries
  - Enforcement of laws dificullt especially in developing countries
  - Discrepancy between penalties, caused damage and additional workload and profit
  - Supposed economical interests
- First success stories published
- Increase rate of spam during 2nd half of 2004 lower than before



Sender and infected computers

- Put known senders of spam in blacklists
  - blacklists for email addresses (inefficient)
  - blacklists for computers (IP addresses)
  - blacklists for advertisement URL's in spam
- Entry in blacklists is automated
- Timely Reaction in response to spam attacks
  - Real Time Blacklists (RBL)
- Queries usually done using the DNS protocol
- Blacklists used in Spamassassin (see later)
  - X-Spam-Report: RCVD_IN_SORBS_DUL,URIBL_SBL,URIBL_WS_SURBL

Some Blacklists

- Spamhaus (SBL)
  - The Spamhaus Block List (SBL) is a realtime database of IP addresses of spam-sources, including known spammers, spam gangs, spam operations and spam support services. The database is kept updated around the clock by Spamhaus Project workers around the world.
- Spam and Open Relay Blocking System (SORBS)
  - It was felt that by publicising a list of compromised hosts, the ever-increasing flow of spam through those hosts could be stopped. On the 6th January 2003 the SORBS DNSbl was officially launched to the public.
- Dynamic User/Host List Database (DUHL)
- Spam URI Realtime Blocklists (SURBL)
  - SURBLs differ from most other RBLs in that they're used to detect spam based on message body URIs (usually web sites). Unlike most other RBLs, SURBLs are not used to block spam senders. Instead they allow you to block messages that have spam domains which occur in message bodies.

Problems in using Blacklists

- “Innocent” sites get listed on RBL's
  - Recent example affecting us: Potsdam university
- Remedy: no yes/no decisions for spaminess (see later), use several distinct criteria
- Entries have to be removed manually
  - Maintenance costs for sites hosting RBL's
- Countermeasures of the spammers:
  - Registration of new hosts in a domain (fast)
  - Registration of new domains (expensive)
  - Use of infected computers as a mail relay

Measures against mail relays

- Only for DESY computers possible
- Informing about computer security, applying patches, updates etc. is essential
- Blocking of incoming connection attempts
  - Acceptance of Mails only on known mail servers
    - At DESY since some time enforced (exceptions)
- Blocking of outgoing connection attempts
  - Sending of mails from arbitrary computers only within DESY, **Port 25 blocked to outside** (RSR policy)
  - In Hamburg enforced, **in Zeuthen from 1. 12. on**

Blocking Spam at the Firewall

- Limiting incoming and outgoing mail traffic to very few well maintained computers
  - A short list of exceptions is held on the firewall
  - Previously nearly each computer was acting as a mail server
  - Today 2 (legal) mail servers in Zeuthen
- Blocking of spam senders (IP addresses) is possible, but usually inneficient (better use RBL's)

Effects of blocking port 25

- Single DESY computers or the whole domain do rarely or never get blacklisted
- Limiting effects of infected computers
  - Especially affecting computers outside the domain
- Also blocking of legitimate traffic
  - Sending from mail clients running inside DESY to provider (outside DESY) no longer possible
- Solution: use of the mail client of the providers (e.g. web based mail client)
- Justified exceptions are allowed
  - A short list of exceptions is held on the firewall

Mail server protection

- Spammers could turn down mail service (DDoS)
  - By huge number of connection requests
  - By actually sending a high number of mails. Generates high load on the virus scanner and may delay mail delivery)
- Several load reduction mechanisms
  - Limiting number of simultaneous connections (was active in Zeuthen until Nov, 22)
  - Limiting the server load (load < 3)
    - If higher, temporary failure (451 – please try later)
  - Connection rate throttling
    - Delays when too many connection attempts (10/s)

Greylisting

- Delayed accepting of mail acts as spam filter
  - Spammers usually do send mails only once
  - Regular mail servers try for a long time
- Make use of this distinct behaviour
- Idea: Connection attempts from unknown senders will be refused the first time “451 – please try later”
  - Accept mail after waiting time of several minutes
  - Register successful delivery in data base (automatic whitelisting)
  - known well behaving domains (HEP) get whitelisted

Greylisting (2)

- In Zeuthen test installation working
  - Test candidates need to get registered for greylisting
  - Can be enabled/disabled for individual addresses
- For 1st test user full success
  - Further spam reduction by a factor 5..10
  - Alle “good” mails did arrive
- Risks
  - Delivery of few good mails 5..30 minutes delayed
  - Some mailing lists do try to send only once(ezmlm)
- Does not work for forwarded spam from well behaving sites (need for a DESY site policy)

Filtering during reception of mail

- Exploitation of the SMTP Protocol
  - Receipt of the complete email and scoring of its content
  - **Thereafter** denial of connection depending on scoring still possible
  - Mail is formally not delivered
- Advantages:
  - Sender does get error receipt on failure
  - Real sender (computer) has to cope with delivery error
  - No delivery error mails to uninvolved users (Mail protocol can easily be faked)

Spamassassin

- Name of the program used to score mails
  - Optimized to not wrongly classify “good” mails
  - De facto standard in Open Source solutions
  - Since Sep, 22 version 3 from Apache available
  - In Zeuthen used since Sep, 27
  - Much improved algorithms compared to 2.6x
- Main task: calculation of a spam “score”
  - A value of 5 is a good sign for spam (about 1:200)
  - At a score of 10 extremly large certainty for mail being spam (better than 1:100000)
- Rejection of mails at score > 10 since Sep, 27

A spam mail example

- Sender and IP addresses modified by me

Date: Fri, 05 Nov 2004 04:46:30 +0000

From: jadestar@mcourtney.com

To: Wolfgang.friebel <wolfgang.friebel@desy.de>

Subject: Re[1]:listsoft

http://257.173.170.21:8180/sft/

- Result of scoring

X-Spam-Level: *****+++++++++

X-Spam-Status: MEDIUM ; 59

X-Spam-Report:
BAYES_50,DOMAIN_RATIO,NORMAL_HTTP_TO_IP,NO_REAL_NAME,
RAZOR2_CF_RANGE_51_100,RAZOR2_CHECK,URIBL_SBL,WEIRD_PORT

Scoring algorithms

- Extremly large number of rules (>800, >500 active)
- Use of (free of charge) blacklists
  - SORBS, Spamhaus, SURBL
- Use of (fuzzy) data bases for known mail contents
  - DCC, Razor
- Recognition of mail protocol violations
  - e.g. sending host pretends to be the local host
- Frequency of words in good mail vs. spam
  - Bayes algorithm, needs training, very powerful
- Further methods would justify separate talk...

Judgement of the filtering used

- Strictly speaking no filtering, as sender cannot deliver the mail
- sender is violating several rules simultaneously
  - Gets informed about this fact
- Users (we) get back ability to check spam folders
- Less error prone mail handling (deleting mails by mistake)
- Technically a delivery to special folders instead of mail rejection possible
  - High resource consumption, same effect for recipient (does not care), worse for sender (does not notice)

Mail filtering by the user

- Takes place when mail is stored in INBOX
  - As requested by the user (“spamfilter on”)
  - At this point in time mail is already accepted
- Or in the mail client on usage
  - Requires creation of simple filter rules
  - Documented for pine, mozilla, Outlook (see web pages Computing->Services->Mail)
  - Filters make use of additional headers inserted by spamassassin (X-Spam-Level)
  - X-Spam-Level: *****++ means score 5.2
  - (in Hamburg only integer part is used)

Virus filtering

- Spam filtering is blocking ~50 percent of viruses
- Another 25 percent get scored with a score > 5
- Remaining virus mails do **not** get filtered
  - Technically easily possible: blocking of suspicious MIME attachments and .exe files
  - Superior mechanisms for Windows at DESY in use:
    - On access scan of all files
    - Frequent updates of virus signatures
    - Quarantine mechanisms
- Viruses in mails are only a partial problem
  - would e.g. not tackle problem of malicious URL's

Virus filtering (2)

- For test purposes installed (currently not active)
  - Virus filters for email: amavisd-new, mimedefang
  - Based on McAfee scanner for Windows and Linux
  - Plugins for Spamassassin mit clamav (Open Source)
- Main problems
  - High resource consumption
  - Missing quarantine mechanisms
  - Covers only part of the problems (download of files, whose URL is contained in the mail body)

Phishing

- Sending of mails suggesting to come from business partners
  - Microsoft, Telekom, PayPal, eBay, ...
- Request to send sensitive or personal data
  - PIN's, bank account numbers, credit card data, ...
- Most simple recognition using text based mail readers (**pine**) and looking at headers
- Common sense is most helpful (“brain on”)

Good mails in the spam folder

- False positives cannot be avoided
  - Wrong classifications (e.g. sender in blacklist)
  - Differing algorithms in Hamburg and Zeuthen, signatures (X-Spam-Level) identical
  - Misclassification also triggered by incomplete (can never be complete) filter training
- Early recognition of false positives is of advantage
  - frequent inspections of the spam folder
  - **New:** once per week mail is sent to user containing liste of spam mails sorted by score (spamfilter on)

Summary

- Measures to suppress spam at DESY
  - No world wide publishing of mail addresses
  - Mail server protection (internal and by firewall)
  - Rejection of mails with bad content
  - Greylisting (experimental)
  - Mail tagging using Spamassassin
  - Filtering of Spam (user based)
    - At mail server (spamfilter on)
    - In mail reader
  - If filtering at mail server
    - Weekly notification about spam in spam folder
    - Removal of spam older than 5 weeks

Future plans

- Evaluation and deployment of greylisting for DESY?
- Virus filtering on the Zeuthen mail server
  - User based activation if possible
- Infrastructure for training of the bayes data base
  - Web interface
  - Data bases per user or one combined DESY data base?
- Regular publishing of the filter statistics on the web

Ultimate goal

- A spam free zone