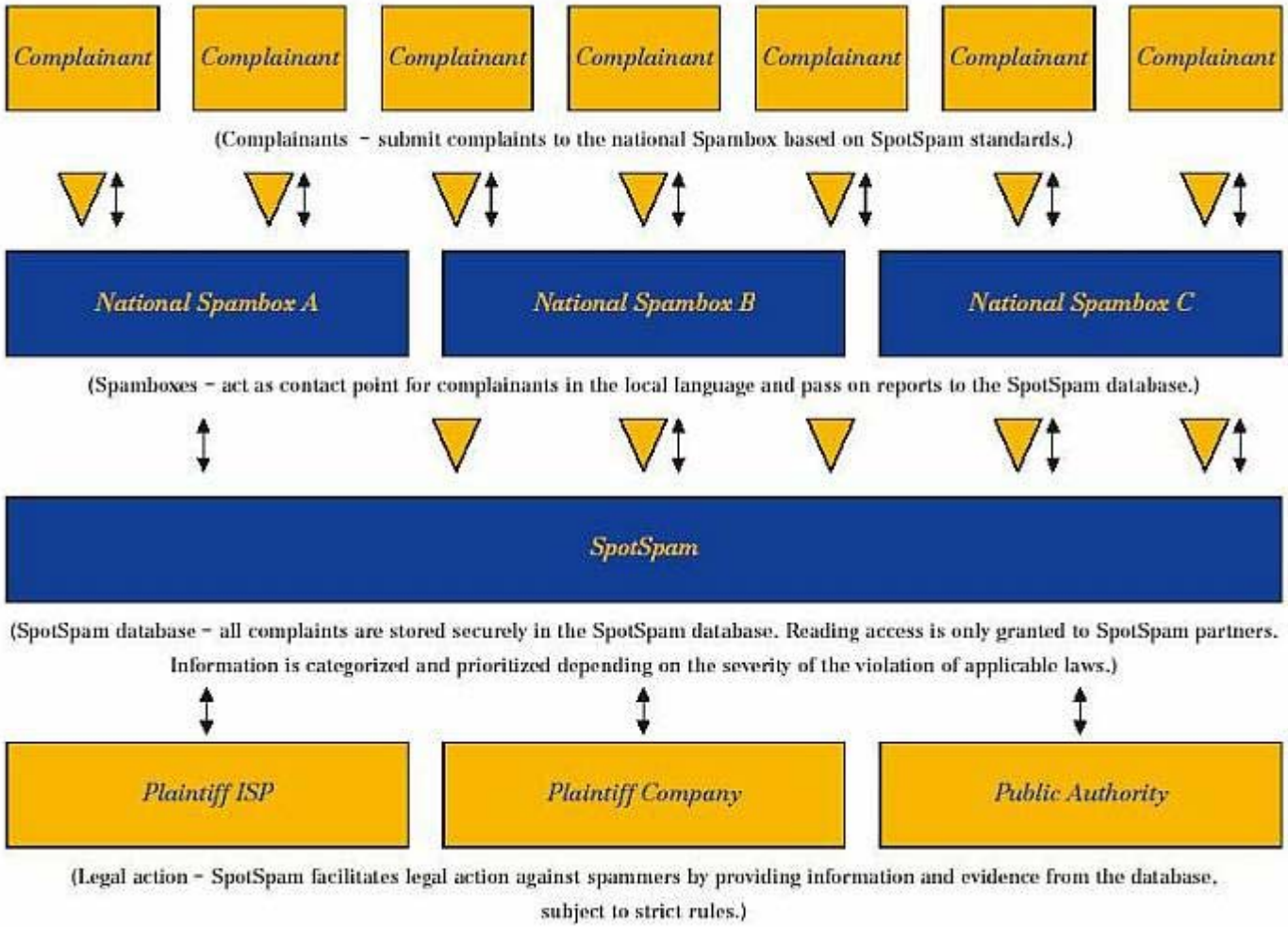
The very last update on the EU project called

SPOTSPAM

Przemek Jaroszewski
CERT Polska / NASK

przemek@cert.pl

- SPOTSPAM is a EU project under EC's Safer Internet Programme
- Consortium: eco (Association of German Internet Providers) & NASK
- Support: Microsoft
- Goals of the project: prepare legal and technical basis for gathering and sharing evidence against spammers
- Timeline: September 05 – September 07


Complainant
Complainant
Complainant
Complainant
Complainant
Complainant
Complainant
(Complainants – submit complaints to the national Spambox based on SpotSpam standards.)
National Spambox A
National Spambox B
National Spambox C
(Spamboxes – act as contact point for complainants in the local language and pass on reports to the SpotSpam database.)
SpotSpam
(SpotSpam database – all complaints are stored securely in the SpotSpam database. Reading access is only granted to SpotSpam partners. Information is categorized and prioritized depending on the severity of the violation of applicable laws.)
Plaintiff ISP
Plaintiff Company
Public Authority
(Legal action – SpotSpam facilitates legal action against spammers by providing information and evidence from the database, subject to strict rules.)

- Spambox operators sign an agreement with SpotSpam
- Complainants register with their local Spambox. They must agree to submit signed evidence in case a court case is launched. They must also certify that all reports submitted by them will actually be spam.
- The reports are stored in central database, which can be queried against IP ranges, email addresses, message subjects etc. Only some indicative data is returned (eg. how many reports are found to match given criteria)
- Interested parties can request full data upon indetification if these are required to launch a court case

- From each reported message the following set of information is extracted:
  - individual attachments
  - IP addresses and associated whois data
  - e-mail addresses
  - spamvertized URLs, associated domain, IP(s) and whois data
- Messages are clustered into spam campaigns according to occurence of similar strings (calculated with Rabin's fingerprints) and identicality of attachments
- Spam campaigns are classified with Naive Bayes algorithm for better prioritization
- Lots of information about IP addresses, domain addresses and their relations is collected, including whois information and geolocalisation

Currently input can be accepted from:

- The prototype spambox application
- Unix mailbox files (bulk submission, mainly for testing)
- HTTP POST request (preferred for external cooperation)
- Email forwarding

The database can be queried against several fields:

- Subject contents
- IP addresses
- Email addresses
- URLs

Full text search is not really a good option for a very large database. As a better (read: very fast and reasonably powerful) solution, indexed hashes of most popular strings will be exercised.

An external partner can only retrieve indicative numerical values while the operators are presented with full set of messages/campaigns that fit the criteria.

- The operator has access to all information about messages related to given URLs, emails or IP addresses.
- Data about misused URLs and IP addresses is periodically extracted from the database, mapped, and can be distributed to external partners (note: it does not include any information about reporters, message contents etc.)
- Complete report covering all data about a given campaign can be generated in pdf format. Such a report can be provided upon verified request for data.

- Over 20 MoUs were signed by parties interested in participation
- SpotSpam was presented in 5th German Anti-Spam Summit (DASK) with over-enthusiastic reception
- The database facility will be mantained as a prototype
- EU support ends in few days
- We need some (reasonably limited) information for exchange
- We need success stories to find ways of financing further maintanance and development
- Possible branch: a standalone reporting/analysis tool with a capability to exchange information with other instances of the same software and beyond

- **More information can be obtained from:**
  - http://www.spotspam.net/
  - mail@spotspam.net
  - myself in person or by email: przemek@cert.pl

CERT POLSKA
zgłaszanie incydentów: cert@cert.pl
strona internetowa: www.cert.pl
tel. +48 (22) 523 12 74
fax. +48 (22) 523 13 99
adres pocztowy:
NASK – CERT Polska
ul. Wąwozowa 18
02-786 Warszawa
Polska

DZIĘKUJEMY ZA UWAGĘ

ZMYSŁ TELEKOMUNIKACJI
NASK

CERT POLSKA