HoneySpider Network

Fighting client side threats

Piotr Kijewski (NASK/CERT Polska)
Carol Overes (GOVCERT.NL)
Rogier Spoor (SURFnet)

20th Annual FIRST Conference on
Computer Security Incident Handling,
June 22-27, Vancouver

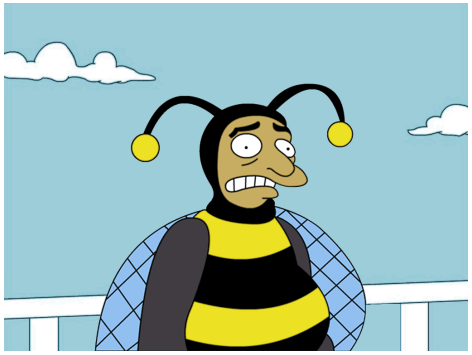
Goals

- **Introduction honeyclients & malicious servers**
- **Technical ins and outs HoneySpider Network**

Outline

- Honeyclients
- Malicious servers
- HoneySpider Network – Why ?
- Project status
- Technical concept
- Wrap up

What is a Honeyclient ? (I)



Definition:

Honeyclients are active security devices in search of malicious servers that attack clients. The honeyclient poses as a client and interacts with the server to examine whether an attack has occurred.

Source:
http://en.wikipedia.org/wiki/Client_honeypot_/_honeyclient

What is a Honeyclient ? (II)

Different honeyclients depending on level of interaction:

4. Low interaction honeyclients
5. High interaction honeyclients

Low Interaction Honeyclient

- Light weight or simulated clients (web crawler)
- Identifies known attacks based on:
  - Static analyses
  - Signatures
- May fail to emulate vulnerabilities in client applications
- Tools:
  - HoneyC
  - SpyBye
  - PhoneyC

High Interaction Honeyclient

- Fully functional operating system with vulnerable applications (browsers, plugins)
- Detection of known/unknown attacks via comparison of different states (before and after visit of a server)
- Slow & prone to detection evasion
- Tools:
  - Capture-HPC
  - MITRE Honeyclient
  - HoneyMonkey

Malicious servers (I)

- **Drive-by download**
  - Download of malware without knowledge of the user
  - Malware offered and executed through exploitation of (multiple) vulnerabilities in a browser, plugin, etc
  - Specific targeted based on browser (IE/Firefox), JVM versions, patch level operating system

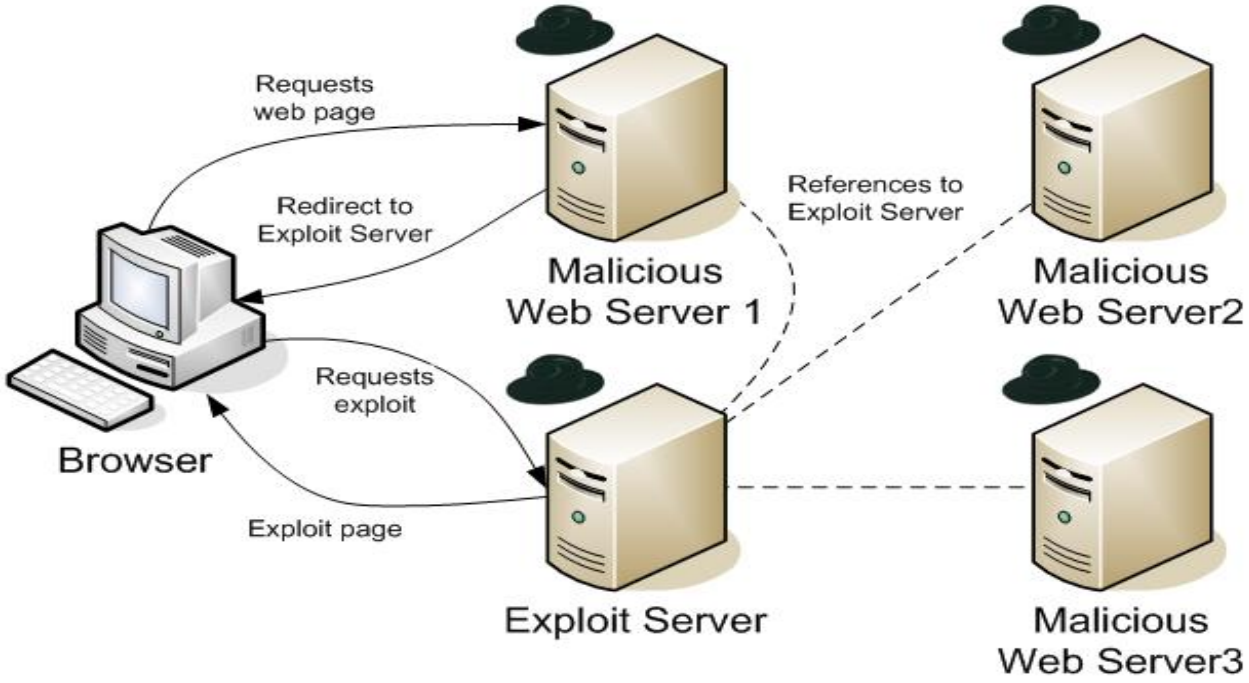
Malicious servers (II)

- **Code obfuscation**
  - Hide the exploit-vector
  - Evasion of signature-based detection (AV products, Intrusion Detection Systems)
  - Examples seen for Javascript, VBScript

```
function xor_str(plain_str, xor_key){
   var xored_str = "";
   for (var i = 0 ; i < plain_str.length; ++i)
      xored_str += String.fromCharCode(xor_key ^ plain_str.charCodeAt(i));
   return xored_str;
}
var plain_str =
"\xf6\xdb\xdc\xdb\xdc\xa0\xb7\xa4....\xff\xed\xdb\xdc\xdb\xdc";
var xored_str = xor_str(plain_str, 214);
eval(xored_str);
```

Malicious servers (III)

Exploits imported from other servers via iframes, redirects, Javascript client side redirects



Source:
http://www.honeynet.org/papers/mws/KYE-Malicious_Web_Servers.htm

Honeyclient project – Why?

- Number of browser exploits increased last years
- Better understanding client side threats
- Existing tools lack in:
  - Integration & management
  - Stability & maturity
  - Limited heuristics
  - Stealth technology
  - Self-learning
- Provide a service to constituents/customers

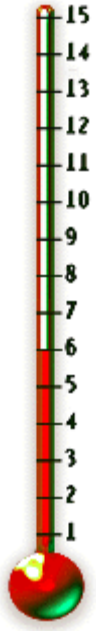
Goal

- Detect, identify and describe threats that infect computers through Web browser technology, such as:

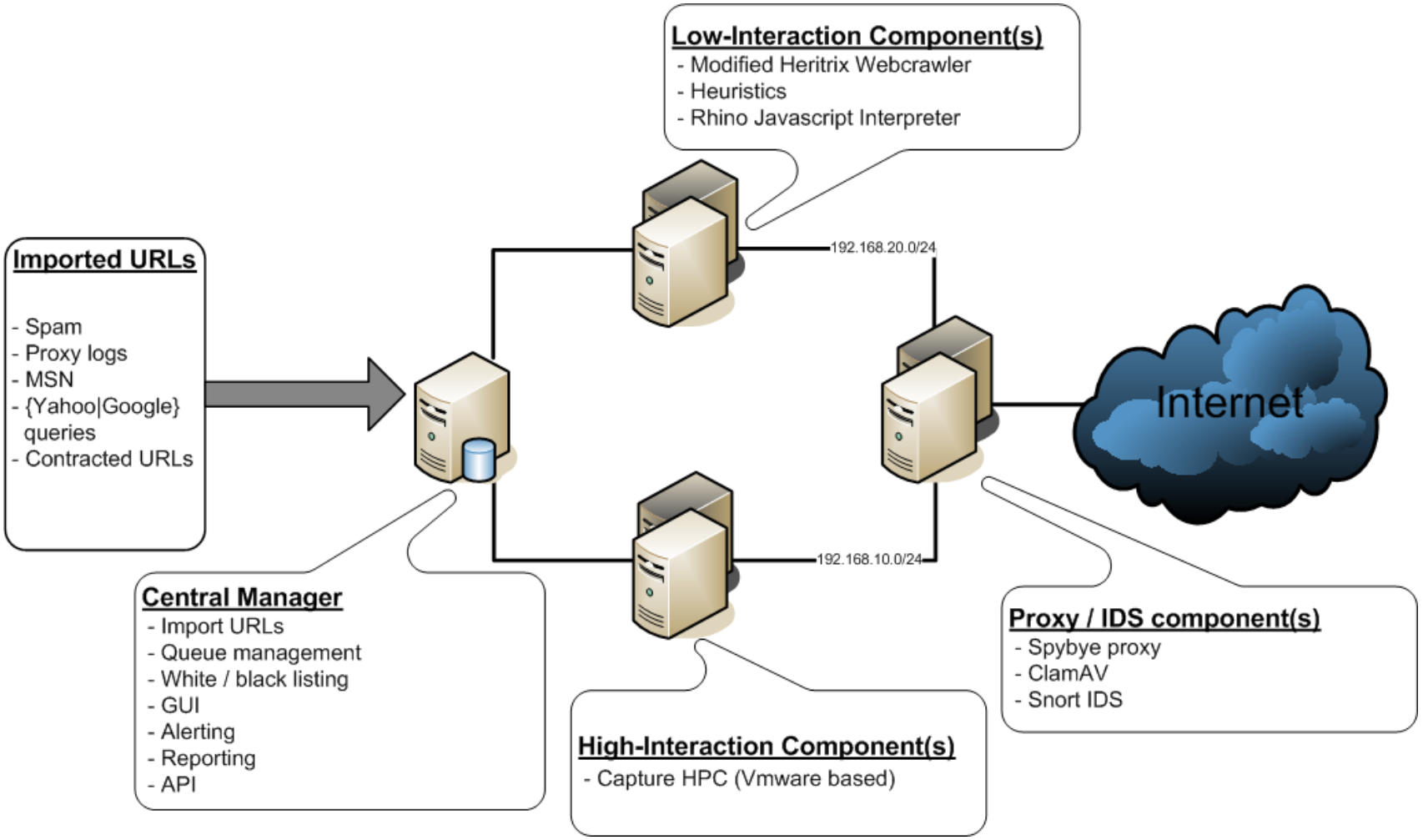
  - Browser (0)-day exploits
  - Malware offered via drive-by-downloads

Project status

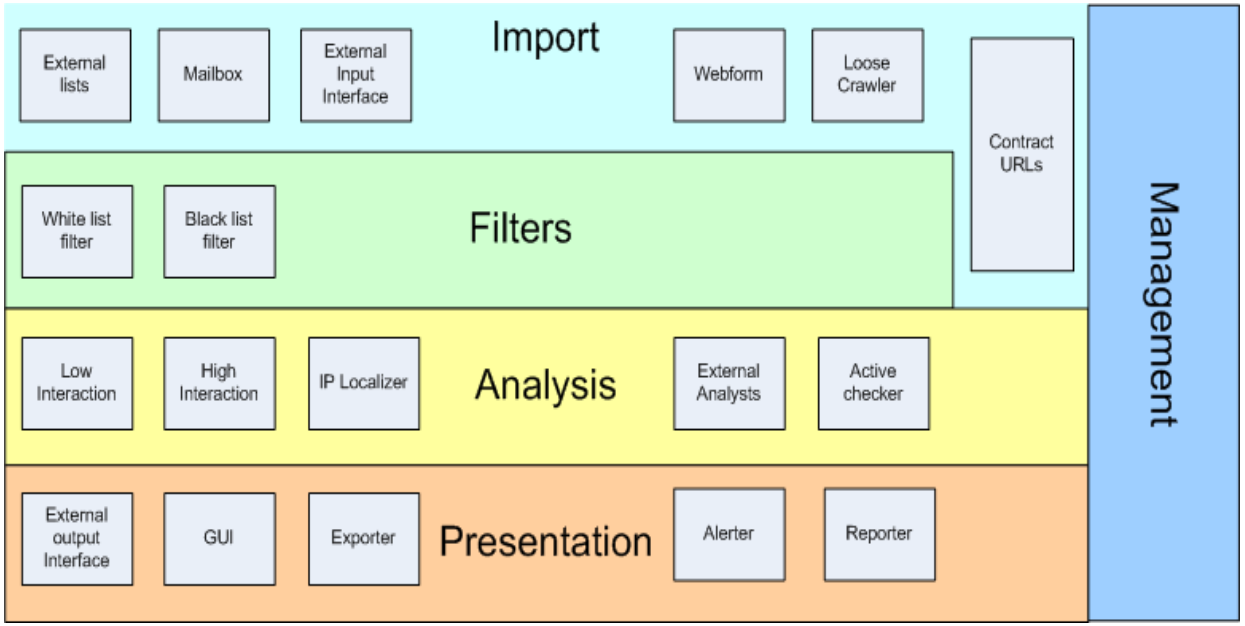
- Completed functional & technical requirements
- Organized project management
- Frequent meetings face-2-face & videoconference
- Started software development September 2007
- 1st Milestone of software developed & currently tested
- Development 2nd Milestone started
- Project will be finished first quarter 2009



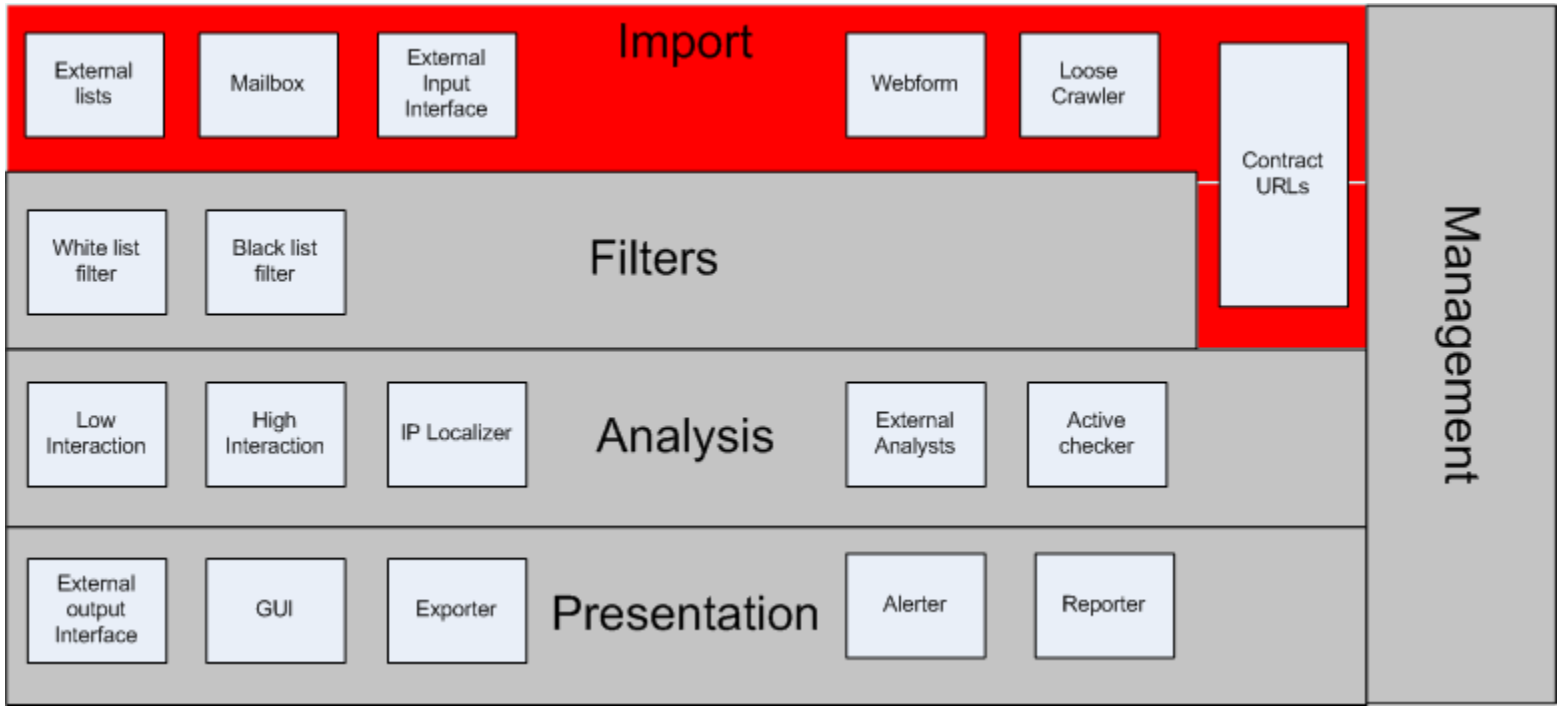
Architecture



Technical concept



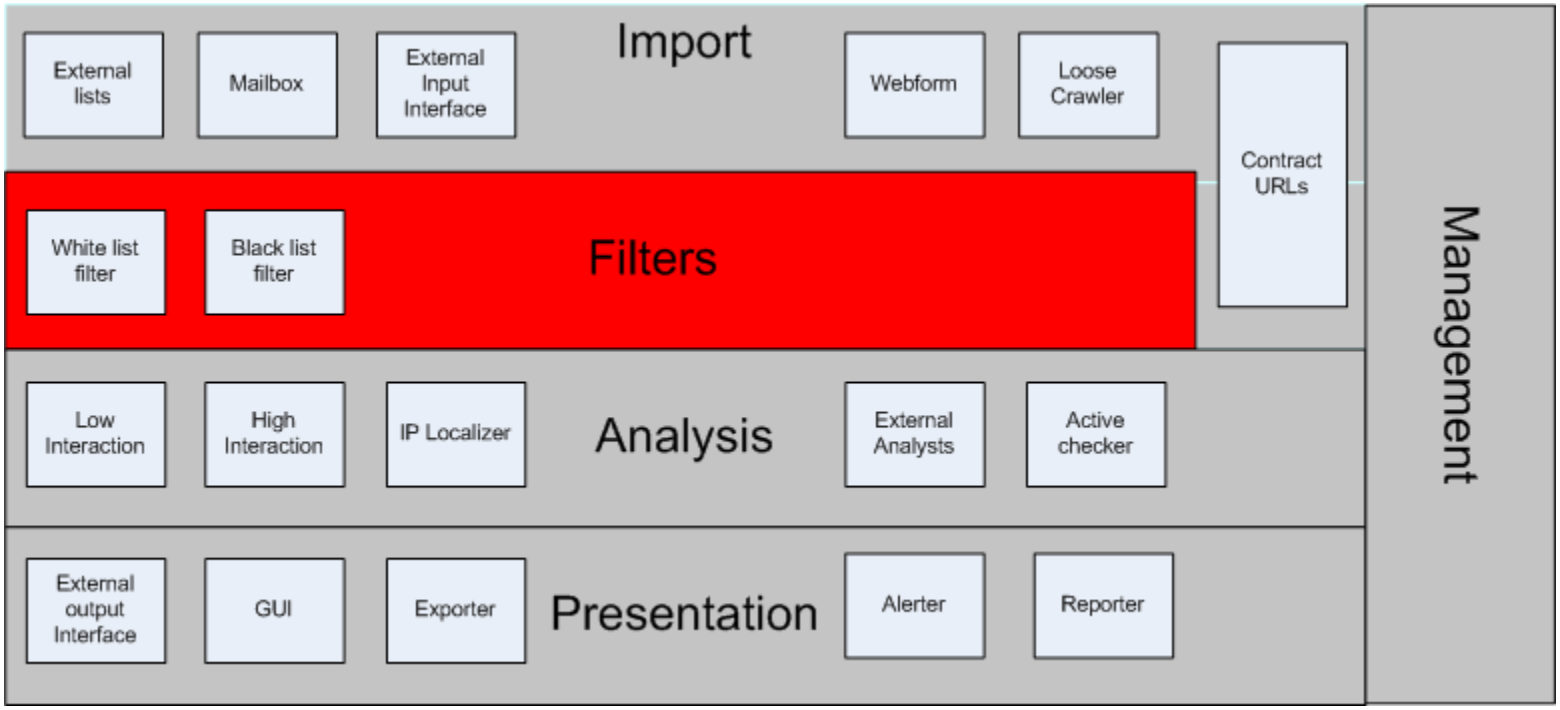
Import layer



Import layer

- URLs (aka objects) report to the import layer via agents (scripts)
- URLs prioritized depending on importance / origin (configurable)
- Contracted URLs:
  - Important URLs which need to be checked frequently (sites of constituents / customers)
- Web form:
  - Manual submission of URLs
- Loose crawler:
  - URLs from {Google|Yahoo}-queries

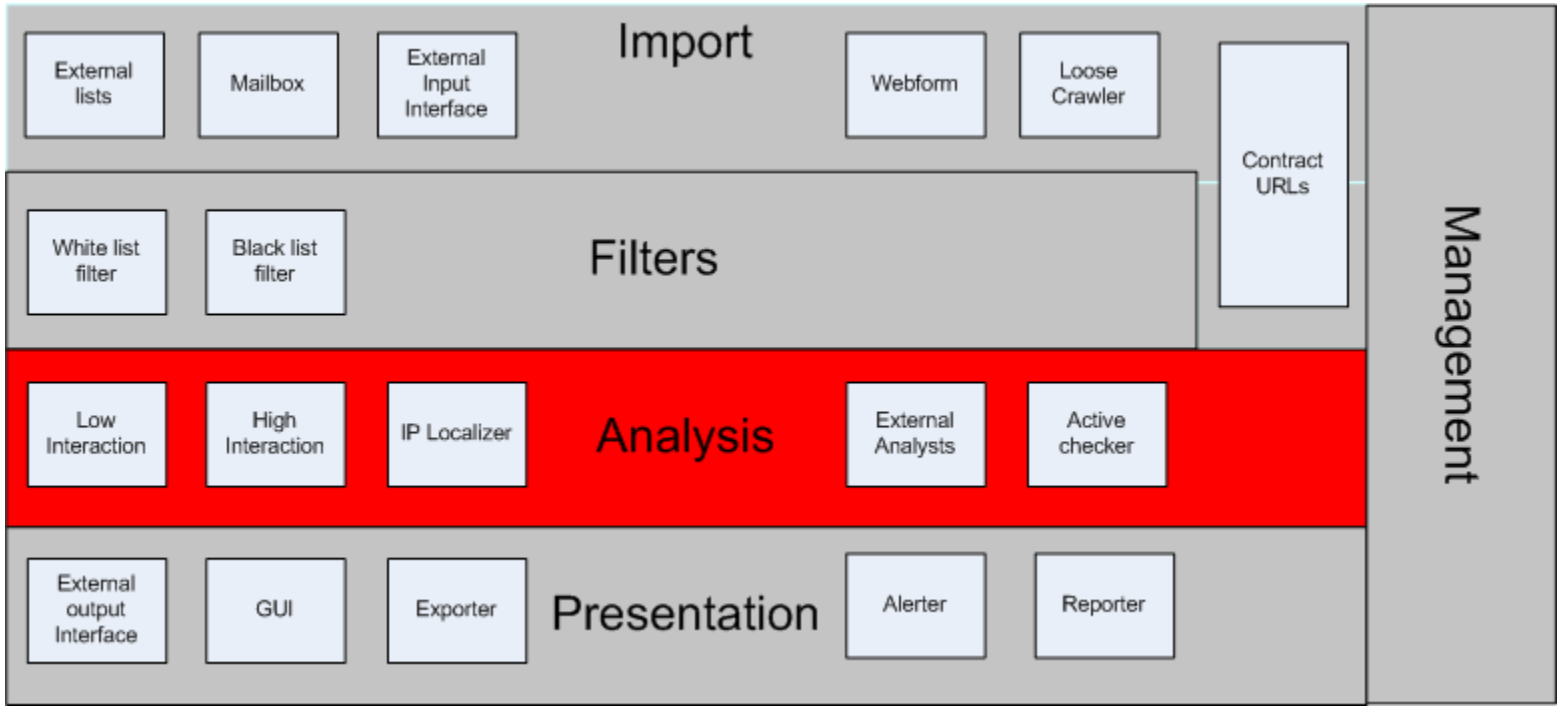
Filter layer



Filter layer

- Filter URLs which are:
  - Already analyzed
  - Not active (domain or IP unreachable)
- Applies on URLs from every source, except contracted URLs
- Black list filter:
  - URLs identified as malicious
  - Hit count & TTL on URL
- White list filter:
  - URLs identified as benign
  - Hit count & TTL on URL (or permanent listed)

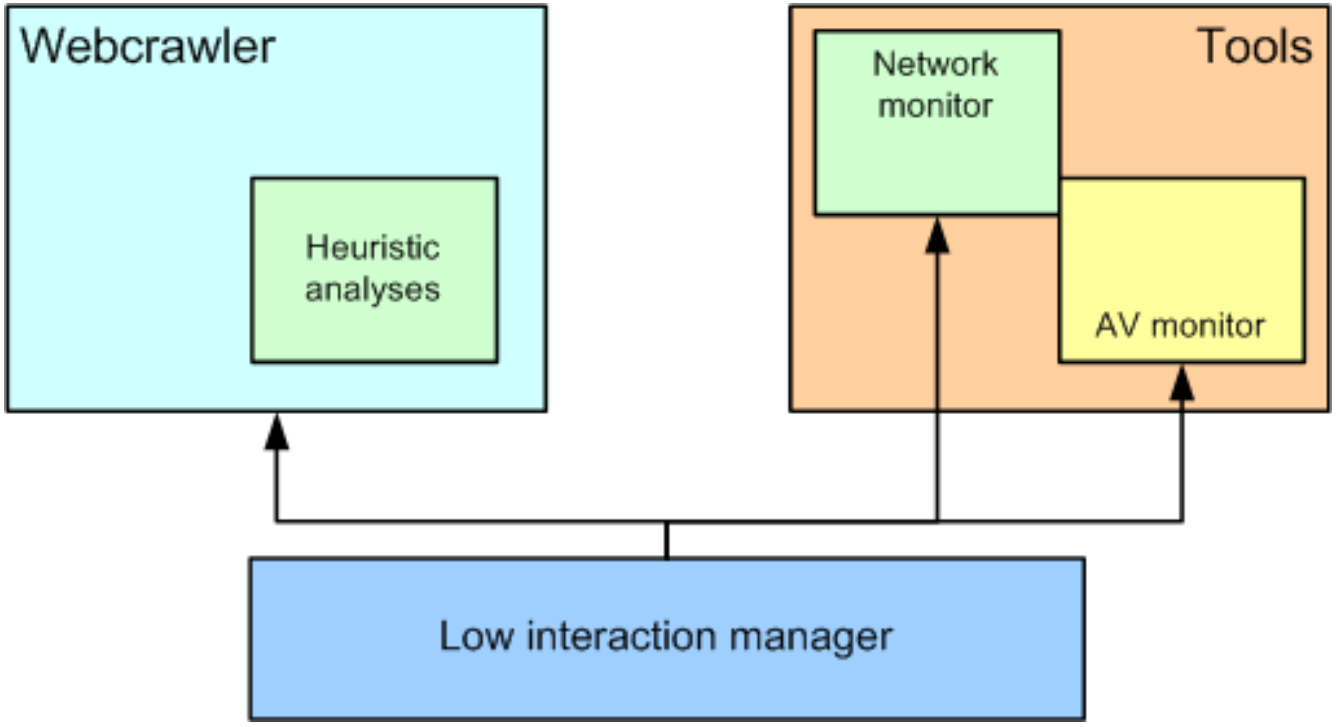
Analysis layer



Low interaction component (I)

- Webcrawler (Heritrix)
- Proxy (Spybye) with ClamAV
- Snort IDS
- Pcap dumps
- Extensions:
  - Rhino (JavaScript engine)
    -> Javascript de-obfuscation
  - Heuristics
    -> Identify obfuscated & malicious JavaScripts

Low interaction component (II)



Low interaction component (III)

- **Heuristics**
  Currently used to identify obfuscated JavaScripts. In the future also used to identify obfuscated VBScripts and to classify websites (*benign, suspicious, malicious*).

- **Current implemented heuristics**
  - Weka Classifiers (machine learning techniques)
  - JSAdvancedEngineDetection
  - JSIterationCounter
  - JSExecutionTimeout
  - JSOutOfMemoryError

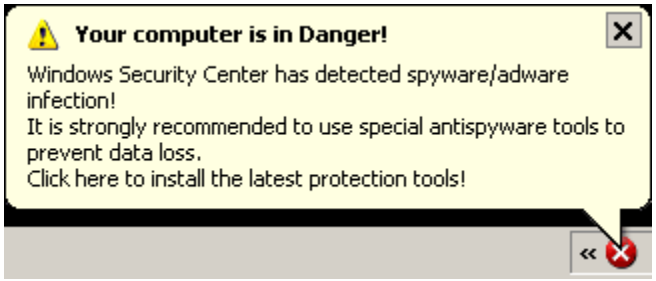
Low interaction component (IV)

- **Heuristics under research**
  Detect malicious web content the same way as detection of spam.

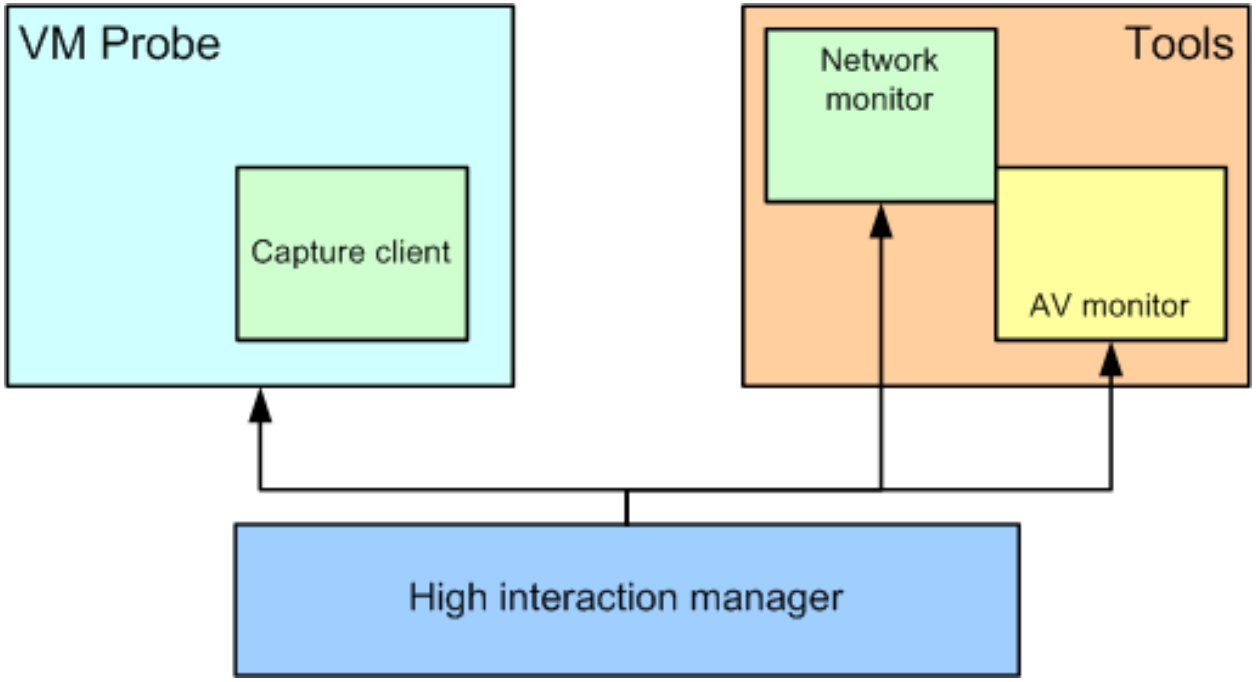
- **Most promising heuristics**
  - Naïve Bayes
  (good test results, undergoing further testing 'in the wild')

High interaction component (I)

- Based on Capture-HPC
- Multiple patch levels Microsoft Windows
- IE / Firefox (possibly plugins, like QuickTime & Flash)
- Checks for:
  - Started or terminated processes
  - Filesystem modifications
  - Registry modifications
- Proxy (Spybye) with ClamAV
- Snort IDS
- Pcap dumps



High interaction component (II)



External analysis

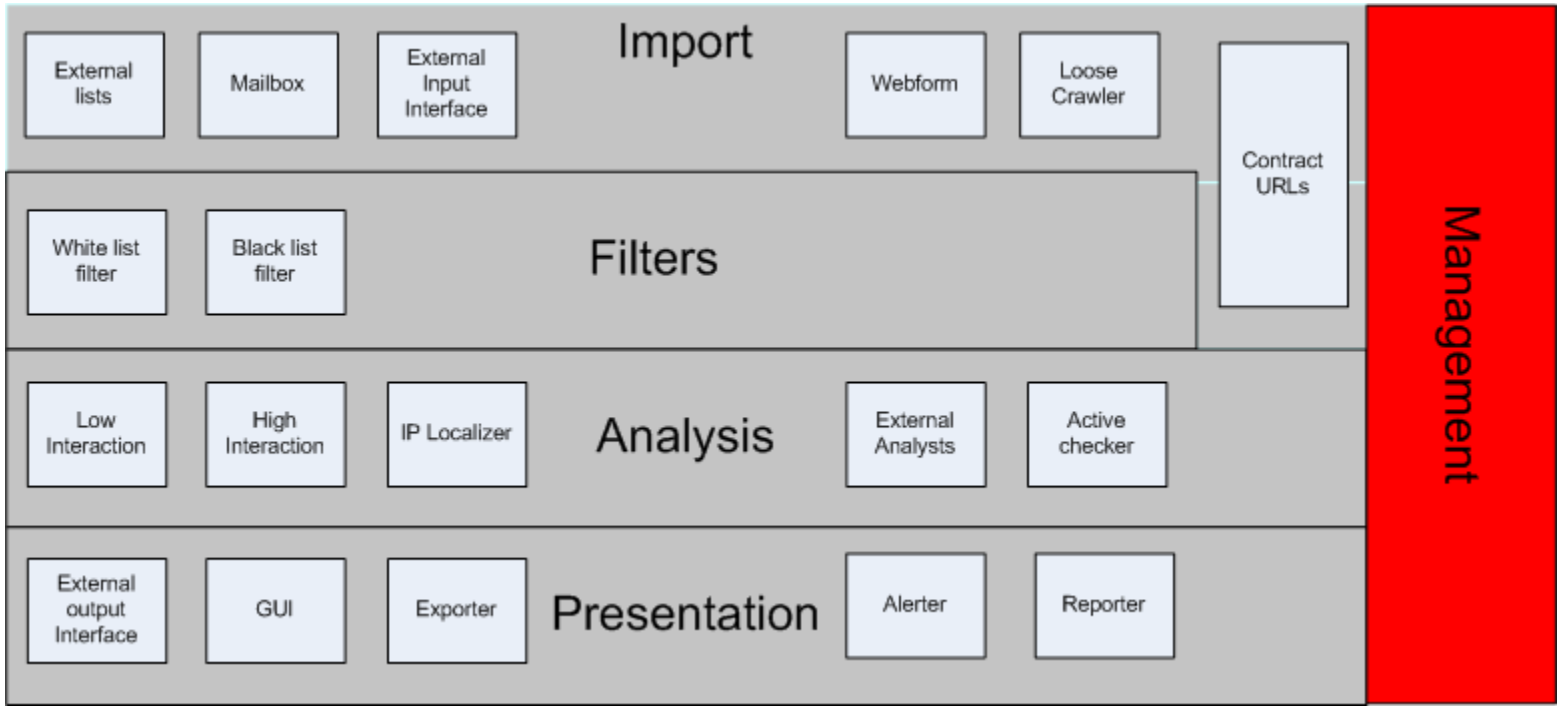
- Submission of a binary file or URL to external sources
- Results are stored in a database
- Plugins for:
  - VirusTotal
  - Anubis
  - Norman Sandbox
  - CW Sandbox
  - Stopbadware

And more analysis…

- **URL Localizer**
  - ASN
  - Name of the ISP
  - Country

- **Active checker**
  - Check if domain still resolves
  - Check if server is active

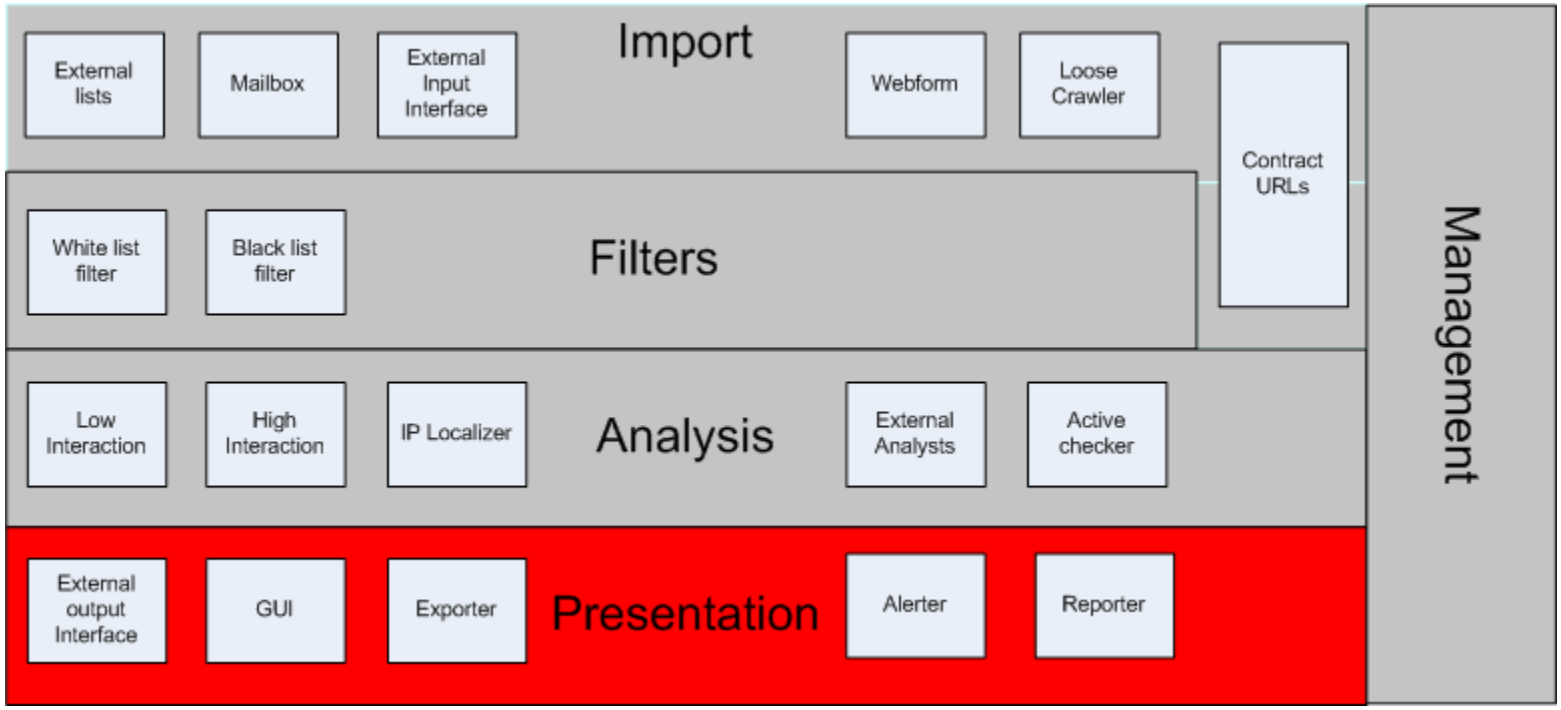
Management layer



Management layer

- **Objects tagging**
  - Confidence level
  - Priority level
  - Process classification
  - Alert classification
- **Queue manager**
  - Manages the main object-queue
- **Signature manager**
  - Generation of signatures
  - Judge quality of signatures
  - Distribute signatures to {Network|AV} monitor

Presentation layer



Presentation layer

- **Web-based GUI**
- **Alerter plugin**
  - Sends alerts via email, SMS
- **Reporter plugin**
  - Creates reports (PDF) with graphical statistics and/or detailed information
- **External output plugin**
  - External systems can fetch results of processed objects

Wrap up (I)

Honeyclients

- ✓ Honeyclients are active security devices in search of malicious servers that attack clients
- ✓ Low-interaction honeyclient currently used to detect known attacks
- ✓ High-interaction honeyclient used to detect known & unknown attacks

Wrap up (II)

Honeyclient project

- ✓ To identify suspicious and malicious URLs
- ✓ A combination of low- & high-interaction honeyclients
- ✓ Many URLs from multiple sources processed based on importance

Links

- HoneySpider Network
  http://www.honeyspider.org/
- Capture HPC
  https://projects.honeynet.org/capture-hpc/
- Heritrix
  http://crawler.archive.org/
- Weka
  http://www.cs.waikato.ac.nz/ml/weka/

Questions ?