

**CONTENT**

Background..... 1

Terminology of Spam Detection...2

Technologies of Spam Detection..3

Flow of eMail Scanning.....7

The Network Box eMail Envelope Scanner.....7

The Network Box eMail Message Scanner.....8

The Network Box Anti-Spam System.....9

Network Box Option For Spam.. 12

Anti-Spam Configuration..... 13

Conclusion..... 13

NOVEMBER 2007

*No part of this publication including text, examples, or illustrations may be reproduced, transmitted, or translated in any form or by any means, electronic, mechanical, manual, optical or otherwise, for any purpose, without prior written permission of Network Box Corporation Limited.*

Network Box Corporation Limited,  
 16th Floor, Metro Loft,  
 38 Kwai Hei Street, Kwai Chung,  
 Kowloon, Hong Kong  
 Telephone: +852 2736-2083  
 Fax: +852 2736-2778  
[www.network-box.com](http://www.network-box.com)

**Background**

More than 2,000 years ago, Sun Tzu wrote “if you know yourself but not the enemy, for every victory gained you will also suffer a defeat”. Before you can effectively protect against a threat, you must understand that threat.

So, what is spam? The dictionary definition is “unsolicited junk email – generally advertising some product sent wide-scale to a mailing list or newsgroup”. But that is just a broad definition, open to interpretation; one man's spam is another man's newsletter.

If you ask an end-user, they will often define spam as “mail I don't want” – very precise from that particular user's point of view, but not implementable in a shared computerized system.

If you ask a lawyer, they'll point to emerging legislation such as the USA's CAN-SPAM Act of 2003 – which runs to 21 pages of legalese to try to define the act of sending spam (and will still have to rely on cases and precedents over the coming few years, to further clarify its scope of enforcement).

However, over the past few years, a consensus has begun to be reached, internationally, over what can be defined as spam and users should be protected against. Spam should have all the following attributes:

1. The message is commercial in nature (note that even messages advertising non-commercial offerings, but sent by a commercial entity, are considered commercial in nature).
2. The message is sent without the addressee's consent. The sender of the message should obtain verified permission from the email recipient prior to transmission.
3. The message does not contain clear and accurate identification of the sender (including messages that fail to provide a valid return, operational, email address or the headers of the email messages are altered to mask the identity of the sender).
4. The message does not include a functional unsubscribe facility that will prevent further correspondence with the recipient in a timely manner.

There are essentially two types of addressee consent:

1. The recipient has a pre-existing business relationship with the sender, and has consented to receive messages containing individualized content specific to the recipient's account with the sender.
2. The recipient has undertaken a confirmed opt-in to a mailing list prior to transmission of the message (a confirmed opt-in being defined as a opt-in joining of the mailing list, with an additional step to confirm the accuracy of the recipient's email address).

While executable files can usually be clearly defined as virus-infected or clean, there is a significant gray area concerning spam.

It should be clear, from the above, that while executable files can usually be clearly defined as virus-infected or clean, there is a significant gray area concerning spam. The above definitions do, however, provide a working test for whether a message is spam or not, and lay the foundation for implementing effective protection against this nuisance.

## Terminology for Spam Detection

Spam detection involves being given an email message, and making a determination if that message is spam or not (in this paper, we refer to messages that are not spam as "ham"). If an email message is determined to be "ham", when it is actually "spam", we call this a false-negative (or a "missed spam"). If an email message is determined to be "spam", when it is actually "ham", we call this a false-positive. The goal of any spam detection system is to maximize the rate of successful detection of spam, while minimizing the false-positive rate.

The false positive rate is defined as the percentage of "ham" messages incorrectly determined to be spam, and is often represented as a figure such as "1 in 10,000" or "1 in 100,000" (which correspond to false-positive rates of 0.01% and 0.001% respectively).

The success rate is defined as the percentage of "spam" messages successfully determined to be spam, and is often represented as a figure such as 95% or 98% accuracy.

Note that the success rate is sometimes quoted as the percentage of all emails successfully marked as spam, which leads to confusion in the marketplace and incomparable figures. The ratio of spam-to-ham for individual users varies tremendously; so it makes little sense to base your ratios on such a varied baseline.

The ratio of spam-to-ham for individual users varies tremendously; so it makes little sense to base your ratios on such a varied baseline.

Putting this together into an example gives us the following:

*A company receives 10,000 email messages. 6,000 of these are spam and 4,000 are ham. An anti-spam system correctly determines 5,850 of the spams to be spam (missing 150), but incorrectly determines 1 of the hams to be spam.*

- *The success rate is  $5,850/6000 = 97.5\%$*
- *The false-positive rate is  $1/4000 = 0.025\%$*
- *The spam ratio is  $6000/10000 = 60\%$*

## Technologies for Spam Detection

There are several technologies which can be used to detect whether a given message is spam or ham. A summary of these is given here:

### 1. Co-operative Spam Checksums

This technique involved breaking apart a message, and taking cryptographic checksums of each component of the message. If the message is known to be spam, its components can be submitted to a centralized database with such an indication. To test a message, the database is queried to see if one or more checksums are already listed as spam. Such systems can return a “confidence” level (based upon the trustworthiness of individual contributors, and the number of contributors for a particular checksum).

### 2. Signatures and Spam Scoring

Such systems use lists of “signatures” (often small strings of text, or regular expressions) which match aspects of spam messages. Each signature is given a score, and a total score kept for all matching signatures. The higher the total score, the more likely the message is to be spam.

### 3. White lists and Black lists

A list of words/patterns which make a message “ham” can be maintained in a white list. Similarly, words/patterns which make a message “spam” can be maintained in a blacklist. If a message matches such lists, a determination can be made as to whether or not it is likely to be spam.

### 4. Heuristics

By examining message structure, and recognizing certain known vulnerability exploits, tests can be designed to provide hints to a “heuristic” spam determination.

### 5. Real-Time IP Blacklists

The email headers contain a record of all the IP addresses of gateways that an email message has passed through. These IP addresses can be tested against a realtime blacklists of gateways known to either (a) be known sources of spam, (b) be known open-relays (allowing third-party relating of messages), or (c) be known dial-up networks (which, some consider, should not be directly sending out emails). Should the email message originate, or pass through, such systems, it can be determined to be more likely to be spam.

### 6. Real-Time URL Blacklists

A common technique used by spammers is to provide links to their websites inside email messages. Such URLs can be extracted, and tested against a realtime blacklist of URLs known to be used by spammers. Should the email message contain such a URL, it can be determined to be more likely to be spam.

### 7. URL to IP Mapping and Blacklists

The list of URLs, from the message, can be processed through the Internet Domain Name System (DNS) to perform reverse-DNS lookups and derive a list of IP addresses. These IP addresses can then be checked against a real time blacklist, to determine that the message is more likely to be spam.

**8. URL Categorization**

Databases, such as Surf Control, have the ability to return a category for a particular URL. The list of URLs, from a message, can be processed through such databases, to determine a list of URL categories in the email message. Scores can then be applied to certain categories, to determine a message as more likely to be spam (or to be blocked according to policy enforcement).

**9. Domain Age**

Similar to realtime URL blacklists, spammers often register Internet domains, and then immediately use them, finally discarding them after a few weeks. Should the email message contain a recently registered domain name, it can be determined to be more likely to be spam.

**10. Bayesian Filtering**

Statistical (or Bayesian) filters can be used to automatically maintain word/pattern white lists and blacklists, together with statistical probabilities as to whether the given word/pattern makes the message spam/ham; based on being taught from a collection of "spam" and "ham". Subsequent messages can be tested against this database, to determine the probability that the message is spam.

**11. Challenge/Response Systems**

Based on the premise that messages come from addresses that recipients have a pre-existing business relationship with, challenge-response systems enforce such a relationship. They maintain a database of sender + IP address + recipient tuples, to record who is permitted to send to each recipient. Should a message arrive from a previously unknown sender, that sender can be challenged (via email or some other confirmation mechanism) to ensure that the address is not automated. Such systems typically quarantine a message until the sender confirms his identity.

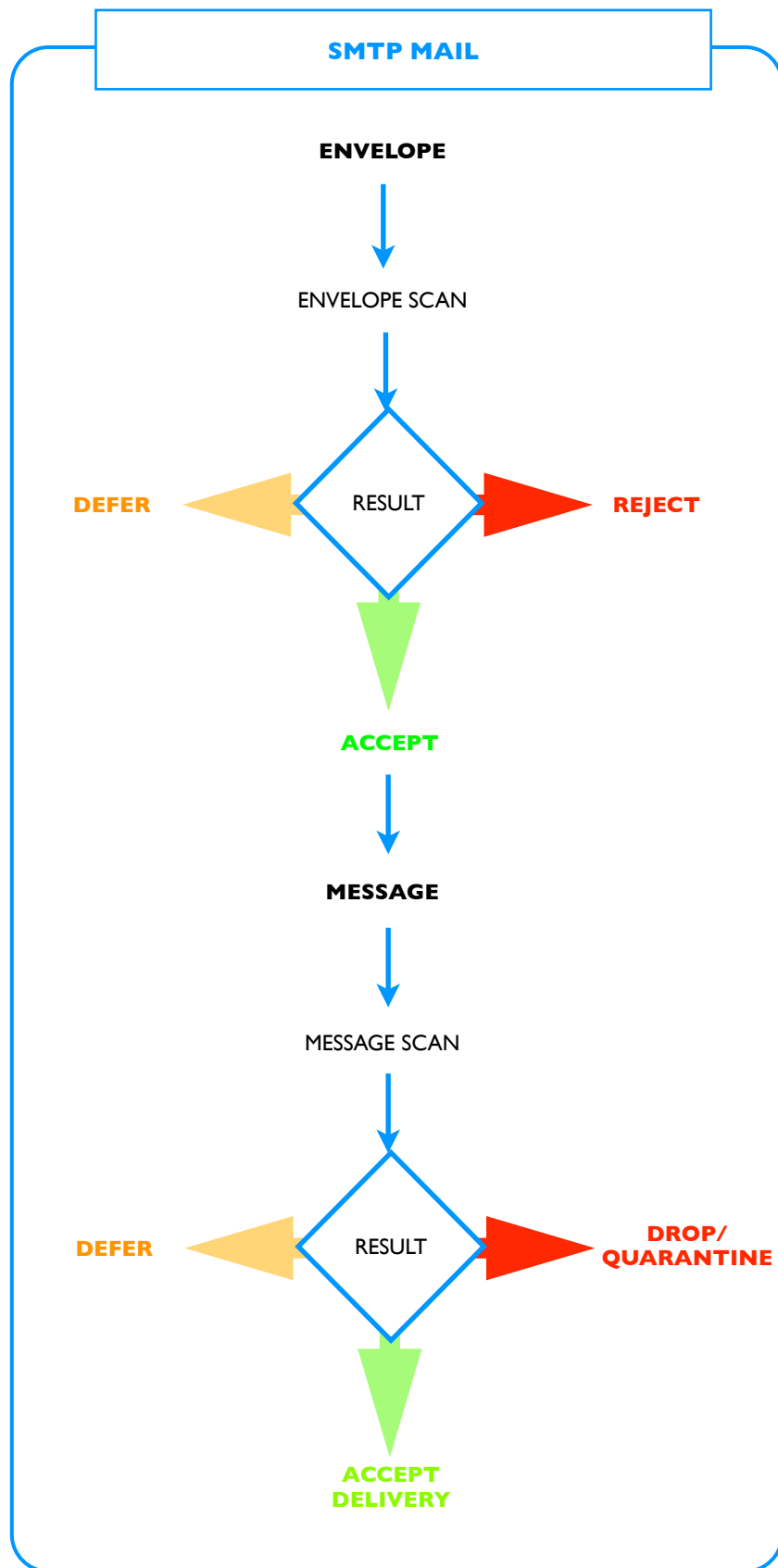
**12. Digital Signatures**

This technique is normally used to indicate that an email message is "ham", and is not used to determine spam, but it can be used to reduce the false-positive rate. The technique relies on the sender calculating a cryptographic digital signature of the entire message, and storing that signature in the headers. The receiver of the message can then determine if the digital signature matches the message itself, to authenticate the sender and message origin, and white list the message as "ham".

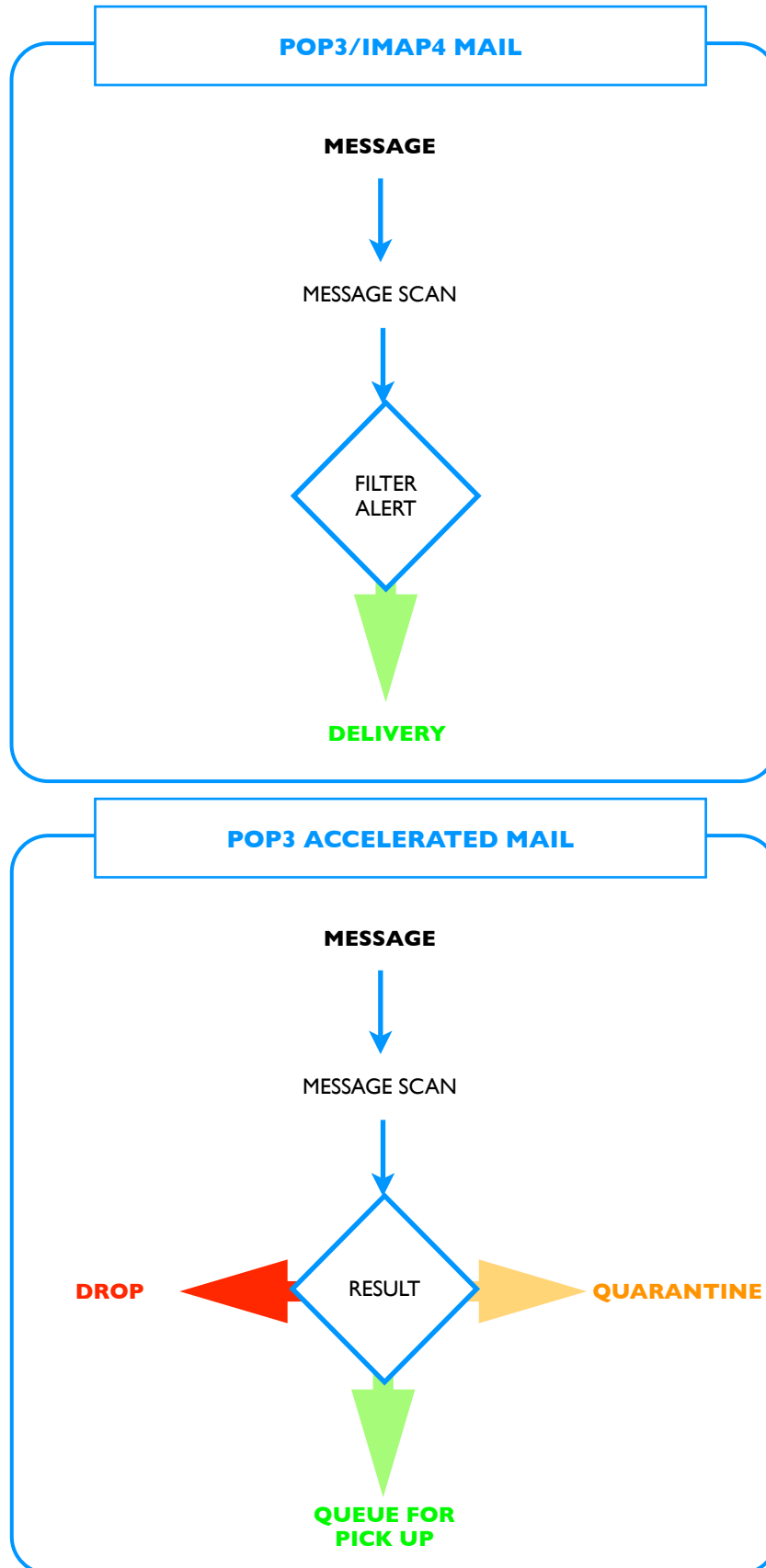
Individually, each of the above techniques is of limited value. But, taken together they form an effective anti-spam system.

Individually, each of the above techniques is of limited value. But, taken together they form an effective anti-spam system.

## FLOW OF EMAIL SCANNING



## FLOW OF EMAIL SCANNING



## Flow of eMail Scanning

There are restrictions inherent in different email protocols, which lead to different capabilities in how they are treated by the Network Box email scanners:

- SMTP email has both an envelope and a message body, and can be quarantined. As such, as the message envelope is received (at the DATA stage of the smtp email transaction) it is passed through an envelope pre-scan and reception of the email message itself is only permitted if this pre-scan is accepted. The full message body is then received and passed through the message scanner.
- Standard POP3 and IMAP4 email messages only contain a message body (no envelope), and can only be filtered (quarantine is not possible, due to protocol restrictions). Such messages are passed through the message scanner and either accepted or filtered (either replaced with an alert message, or marked to indicate they are not acceptable).
- Network Box supports patented technology for the optional acceleration of the POP3 protocol. In such a system the Network Box itself downloads, scans, and stores the POP3 messages ready for the client workstation to collect. While there is still no envelope, this mechanism does permit the quarantining of messages containing malware and spam (which the standard POP3 protocol does not). Such messages are passed through the message scanner and are either quarantined/dropped or queued for collection by the client workstation.

## The Network Box eMail Envelope Scanner

The Network Box email Envelope Scanner operates at the envelope level (containing just message sender, recipients, source IP address and some protocol information). As it operates before reception of the actual message, a block at envelope scan stage can result in significant bandwidth and workload capacity savings.

The system works by scanning the envelope using a variety of engines to determine acceptability. The result of this scan can be one of:

- **Defer** – the message is temporarily deferred (an instruction to the sender to try to re-send the message at a later time).
- **Reject** – the message is permanently rejected (an instruction to the sender to never repeat sending this same message again).
- **Accept** – the envelope is acceptable and the client should continue to send the entire message for further scanning by the Network Box email message scanner.

The Network Box email Envelope Scanner can be optionally connected to customer systems for recipient address verification, and to the Network Box IDP/Firewall system for Directory Harvest Attack protection (and temporary blacklisting of malicious source IP addresses).

## The Network Box eMail Message Scanner

The Network Box Internet Threat Protection system has the ability to comprehensively scan emails for company policy conformance, viruses, intrusions, and spam. Let's examine the flow of email through the scanner:

### eMAIL MESSAGE SCANNER

Initialization



Pre-Scan



Analyze+Unpack



Post-Unpack



Anti-Virus Scan



Anti-Spam Scan



Policy Enforcement Scan



Post-Scan



Alerting

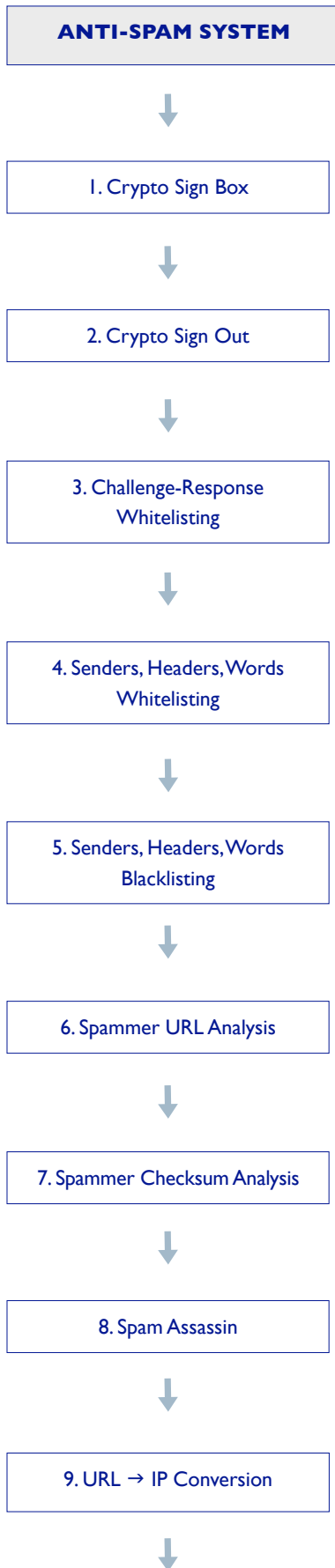


Cleanup

1. Initialization (preparation for the scan)
2. Pre-Scan (cleanup and selection of scanning system)
3. Analyze+Unpack (loop until message completely unpacked and analyzed)
  - Analyze (analysis of email message and embedded content sections)
  - Unpack (unpacking of message structures and attached archive files)
  - At unpack and analyse stage, the Network Box handles extraction of text and images from PDF and office document formats. Images are passed through OCR for text extraction and later analysis.
4. Post-Unpack (cleanup of unpacking system)
5. Anti-Virus Scan (searching for viruses, by heuristics and signatures)
  - Pre-Anti-Virus (preparation of anti-virus engines)
  - Anti-Virus Scan on the Message (scan message body and headers)
  - Anti-Virus Scan on Files (scan attach files)
  - Anti-Virus Scan Content (scan attached content)
  - Post-Anti-Virus (cleanup for anti-virus engines)
6. Anti-Spam Scan (searching for spam, by multiple engines and methods)
  - Pre-Anti-Spam (preparation of anti-spam engines)
  - Anti-Spam Scan on the Message (scan message body and headers)
  - Anti-Spam Scan on Files (scan attached files)
  - Anti-Spam Scan on Content (scan attached content)
  - Post-Anti-Spam (cleanup for anti-spam engines)
7. Policy Enforcement Scan (enforcement of company policy)
  - Pre-Policy (preparation of policy enforcement engines)
  - Policy Scan on the Message (scan message body and headers)
  - Policy Scan on Files (scan attached files)
  - Policy Scan on Content (scan attached content)
  - Post-Policy (cleanup for policy engines)
8. Post-Scan (cleanup for all engines)
9. Alerting (raising alerts)
  - Pre-Alert (preparation for alerting)
  - Alert (issuing of alert messages)
  - Post-Alert (cleanup for alert engines)
10. Cleanup (final cleanup, reporting and logging)

You can see that stage 6 of the scan provides for the anti-spam engines to hook into the scanning system and help decide if a message is spam or not.

## The Network Box Anti-Spam System

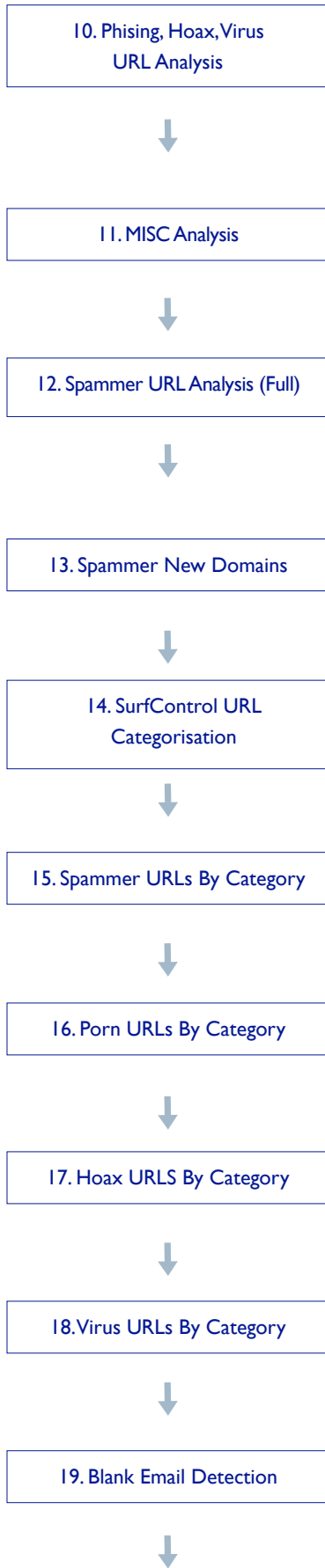


The Network Box anti-spam system currently (August 2005) consists of 24 anti-spam engines, and 175,000 signatures, covering all 12 anti-spam techniques explained in this white paper.

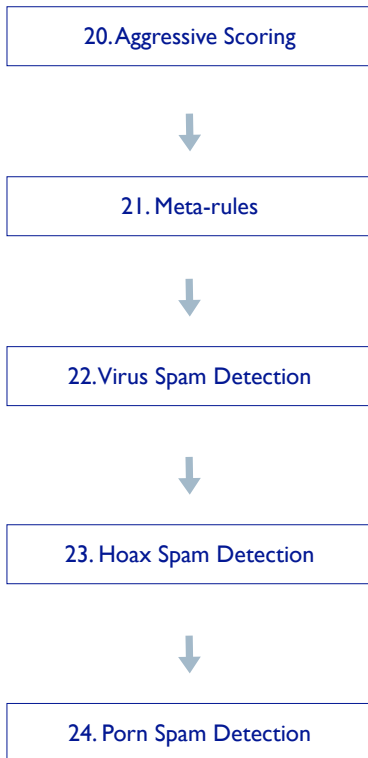
The kernel of the anti-spam system maintains an overall spam “score”, a spam “threshold” (scores above which will be treated as a “spam” result), a whitelist and a blacklist flag.

The anti-spam system hooks into the scanning process at several events, and works as follows:

1. Messages originating from Network Boxes are cryptographically signed. This confirms that the message was (a) sent from a Network Box, and (b) authenticates the sending Network Box identity. The Network Box “as\_whitehamsignbox” anti-spam engine can automatically whitelist such messages.
2. Messages passing outbound through Network Boxes are cryptographically signed. This confirms that the message was (a) sent outbound via a Network Box, and (b) authenticates the gateway Network Box identity. The Network box “as\_whitehamsignout” anti-spam engine can automatically white list such messages.
3. The Network Box “as\_spamcrwhitelist” anti-spam engine recognizes messages from senders who have previously successfully passed a challenge-response challenge, and can automatically whitelist such messages.
4. The Network Box “as\_spamwhitelist” anti-spam engine supports whitelists for senders, headers and words. If an email message is found to match these whitelists, the message is whitelisted.
5. The Network Box “as\_spamblacklist” anti-spam engine supports blacklists for senders, headers and words. If an email message is found to match these blacklists, the message is blacklisted.
6. The analyse and unpack engines can extract a basic list of URLs found in both text and html sections of email messages. These URLs are added to a list stored in the scanning system kernel. The Network Box “as\_spamhinturl” engine compares these URLs against several blacklists (domain, url and IP based), and will raise a score indicating the message as likely to be spam, if found. This is run fairly early on in the scan process, to improve performance by easily blocking messages with known spammer URLs (without having to pass through later, CPU and network intensive, scanning engines).
7. Each component section of the email message is cryptographically checksummed, and the Network Box “as\_spammd5” engine checks such checksums against a known blacklist of spam checksums. If a component of the message matches, the Network Box can be configured to raise a score indicating the message as likely to be spam.



8. The “as\_spamassassin” engine uses the industry standard spamassassin engine to analyze the email message. This engine performs heuristic, signature, realtime blacklist, co-operative checksum and Bayesian analysis, and produces scores which are added to the overall Network Box anti-spam score. Additional URLs found by spamassassin are also added to the list of URLs maintained by Network Box.
9. The Network Box “as\_url2ip” engine uses the Internet Domain Name System (DNS) to convert the list of URLs found in the message into IP addresses. This list is stored in the kernel for later use.
10. The Network Box “as\_spamurl” engine examines the URLs and IP addresses found in the message, and compares this against blacklists for phishing, hoax and viruses. If a match is found, the message is blocked as malicious.
11. The Network Box “as\_spammisc” engine examines miscellaneous attributes of the message (such as phone numbers, email, addresses, fuzzy signatures, etc) and compares against blacklist.
12. The Network Box “as\_spamhinturl” engine is then run again (this time based on the most comprehensive list of url and IP built up so far in the scanning process). The engine compares these URLs against several blacklists (domain, url and IP based), and will raise a score indicating the message as likely to be spam, if found.
13. The Network Box “as\_spamnewdomains” engine runs, to check the registration records of domains in URLs mentioned in the email message. Any domains registered more recently than a pre-defined threshold will raise a score indicating the message as likely to be spam.
14. In co-operation with Surf Control, the Network Box “as\_categorisepolicy” engine then runs to categorize each URL mentioned in the email message. The resulting list of categories is maintained in the scanning system kernel.
15. The Network Box “as\_spamcategories” engine is run the check the category list, and raise spam scores, for each category, as defined in the configuration.
16. The Network Box “as\_porncategories” engine is run the check the category list, and raise pornographic spam scores, as defined in the configuration.
17. The Network Box “as\_hoaxcategories” engine is run the check the category list, and raise hoax spam scores, as defined in the configuration.
18. The Network Box “as\_viruscategories” engine is run the check the category list, and raise virus spam scores, as defined in the configuration.
19. A recent problem is fragmented, blank email messages (contain no email subject or body). The Network Box “as\_spamblank” engine detects such messages and will raise a score indicating the message as likely to be spam, if found.



20. The Network Box “as\_aggressive” engine runs late in the scan process, and can raise the scores (by a configurable factor) to tune the anti-spam system to be more (or less) aggressive for certain types of spam.

21. The Network Box Meta-rules engine combines individual spam rules, by boolean and arithmetic logic, to produce ‘combined’ and other such rules.

22. The Network Box “as\_spamvirus” engine runs to check if the message has had any spam scores raised related to viruses. If found, the engine blocks the message as a virus spam.

23. The Network Box “as\_spamhoax” engine runs to check if the message has had any spam scores raised related to hoaxes. If found, the engine blocks the message as a hoax spam.

24. The Network Box “as\_spamporn” engine runs to check if the message has had any spam scores raised related to pornography. If found, the engine blocks the message as a pornographic spam.

After the message has passed through all anti-spam modules, the system will make its spam determination based on the following rules:

- If the whitelist flag is set, the message is “ham”. Finish analysis.
- If the blacklist flag is set, the message is “spam”. Finish analysis.
- If the final score is greater than or equal to the threshold, the message is “spam”, otherwise the message is “ham”.

## Network Box Option for Spam

eMail messages detected as spam can be treated in several ways (depending on protocol):

FUNCTION	PROTOCOL				
	Standard POP3	Accelerated POP3	Standard IMAP4	Intercepted SMTP	Transparent SMTP
Add an "X-Spam-Status" header to the message	Yes	Yes	Yes	Yes	Yes
Add an "Spam-Check-Result" header to the message	Yes	Yes	Yes	Yes	Yes
Add a prefix to the subject line of the message	Yes	Yes	Yes	Yes	Yes
Quarantine the message (on the Network Box hard disk)	No	Yes	No	Yes	Yes
Drop (silently discard) the message	No	Yes	No	Yes	Yes
Redirect (change the recipient) the message	No	No	No	Yes	Yes

For those spam emails quarantined on the Network Box, the Network Box system can be configured to send daily/weekly summary reports to end-users, and can permit administrator or end-user release of spams from quarantine.

## Anti-Spam Configuration

The Network Box email system is extremely configurable. Individual engines can be enabled/disabled based on tests, including:

- Direction of the message (inbound or outbound)
- Whether the message is being filtered (eg; POP3, IMAP4)
- Whether the message is redirectable (eg; SMTP)
- Globally (ie; for everything)
- Based on the content of a message headers
- Based on the proxy handling the message (eg; SMTP, POP3)
- Based on a single recipient of the message
- Based on a recipient being one of the recipients of the message
- Based on the sender of the messages
- Based on the sender IP address
- Individual engine parameters can be set to adjust scores and weightings to suit end-user requirements.

## Conclusion

The Network Box anti-spam solution is the most comprehensive and effective gateway anti-spam solution in the market today

The Network Box anti-spam solution is the most comprehensive and effective gateway anti-spam solution in the market today. It provides 24 anti-spam engines, combining 11 different techniques and is backed by a database of over 175,000 signatures. It provides true defense-in-depth, in a single managed gateway appliance.

*Network Box Security Response, November 2007*