

# Kaspersky® Anti-Spam 3.0

## Whitepaper

- Collecting spam samples
- The Linguistic Laboratory
- Updates to antispam databases
- Spam filtration servers

# Spam filtration

is more than simply a software program. It is a human and technological process involving:

- 1 The collection of spam samples
- 2 The linguistic laboratory
- 3 Regular database updates
- 4 A filtration server

In this document, we will describe the role of each of these methods in detecting and filtering spam.

# 1 Collecting spam samples

The spam laboratory receives spam from three main sources:

- **Spam traps.** We have a network of spam traps across the Internet, which are registered to pornographic websites, databases of addresses used by spammers and free email services, etc.
- **Members of the public.** We welcome and receive information about spam from members of the public, especially from users of free email services.
- **Feedback from users.** Kaspersky Lab customers and beta testers also send in examples of spam to the company.

Interestingly, each of these sources yields its own genre of spam. Spam tends to vary according to the country from which it originates and how it is positioned on different kinds of mail servers. For example, spam on free email servers such as Hotmail (which is particularly afflicted by offers of university degrees for sale), is of a very different type from that which targets corporate email addresses (mainly invitations to seminars and conferences).

Our spam sources forward all of their incoming mail to an address in our linguistic laboratory.

We currently receive thousands of new spam messages every day (hundreds of thousands of these messages are repeat spam messages that do not have any unique features) and we are constantly adding to our list of sources. Our current forecasts suggest that we will be able to extend our network of spam traps whereby we will detect at least 10,000 new spam messages each day.

## 2 The Linguistic Laboratory

The Kaspersky Lab Linguistic Laboratory employs 12 analysts (and regularly acquires new specialists in foreign languages) and is operational 24 hours a day. All of our spam analysts have advanced qualifications in linguistics and have experience working with applied linguistics and artificial intelligence technologies. We use our own software for analyzing and processing messages.

### Linguistic Analysis

The Linguistic Laboratory receives spam from our spam traps on the Internet. Our linguists scan these messages with our filtration software and weed out new spam messages that have not yet been logged with signatures to databases. These messages are then divided into two groups: legitimate letters (which occasionally also fall into our spam traps) and spam messages under different headings.

The next task is to create a signature for each spam message and add it to the antispam database. These signatures allow our antispam products to detect future mailings of the spam message as well as even slightly adapted versions.

Another tool that helps Kaspersky Anti-Spam 3.0 decide the status of email messages is UDS (the Urgent Detection System). Using this system, the program can request information about new spam mailings in real-time. As soon as a new spam message is discovered, a temporary signature is added to the UDS server, which contains basic information about the message (further details about UDS technology are given in the section, "Spam Filtration Server").

The next task is to create a full spam signature for each message. The linguists conduct a thorough analysis of the messages -- highlighting new phrases, determining the spam weight and adding them to semantic templates. All this prepares the ground for the heuristic analysis stage.

Our specially designed software makes this process fast and highly accurate. By simply highlighting the message, dragging it across and clicking on the arrow, our linguists can determine the spam weight. Tools are provided which allow the linguists to check the quality of detection by:

- a) using new messages (to make sure that detection has improved)
- b) using a sample database containing spam messages (to check that overall detection has not deteriorated)
- c) using a database containing legitimate letters (to check for false positives).

At the same time, our spam analysts analyze each message "envelope", meaning its formal attributes (e.g., sender, recipient, route, etc.), and create new rules for detecting spam based on these attributes.

**The linguistic laboratory works 24/7/365 and updates to databases are added every 20 minutes, while UDS signatures are added in real-time.**

# 3 Updates to antispam databases

Every 20 minutes, semantic templates, message templates and new formal attributes are added to the update server.

Once the filtration server (Kaspersky Anti-Spam 3.0) downloads these updates, it can use the new semantic rules and templates to detect the very newest spam messages.

It is essential that security companies release updates to databases as regularly as possible for the following reasons:

1. Spam lexicon changes very quickly. Although the aims of spammers never change -- e.g., to sell something to recipients, attract them to their websites or force them to send an email response -- the language they use is constantly changing. We have found that when databases are not updated for a week, the quality of spam detection falls from around 90–95% to as low as 40–60% (this is determined by spam lexicon and typical spam attributes).
2. Repeat spam. The same spam messages are often mailed repeatedly, and repeat messages can account for as much as 10 to 15% of messages in users' inboxes. It is not uncommon for the same message to appear five or six times a week, or in some cases, on the same day.
3. Speed of distribution. Obviously, mass mailing 10 million messages requires time. The mail server that has to distribute these messages may be occupied for hours upon hours. Moreover, there are no guarantees that email will be delivered immediately. The last spam letters to be sent can arrive in users' inboxes hours after the first messages have reached their target. Consequently, most users of antispam software receive signature updates for these spam messages before the message reaches their inbox.

# 4 Spam filtration servers

Spam filters are server programs that install on the network gateway and filter spam from incoming email. Kaspersky Anti-Spam 3.0 is a highly scalable solution that provides equally effective and efficient protection for small businesses and large organizations that see hundreds of gigabytes of traffic a day.

The filtration server identifies and filters out unwanted messages in SMTP mail traffic as they arrive on the server and before they reach the end user's inbox.

The main advantages of Kaspersky Anti-Spam 3.0 are:

- **Heuristic and linguistic analysis** of the content of email messages (content filtration);
- **Updates to antispam databases** every 20 minutes;
- **Real-time response** to new spam mailings using UDS technology;
- **Combination of filtration methods** (content filtration and analysis of formal attributes) tightly integrated into one module;
- **Centralized management** of all filtration parameters from a single interface;
- **Scalability** to the requirements of both small businesses and large-scale mail systems.

Kaspersky Anti-Spam 3.0 offers full support for systems working under the Linux and FreeBSD platforms.

## Spam filtration methods used in Kaspersky Anti-Spam 3.0

The filtration server combines content filtration methods with analysis of formal attributes, enabling it to search message content for key words and phrases and compare the message to spam templates in the signature databases.

The following tools are used in the filtration process:

1. **Lists.** Sender IP addresses are checked against blacklists of spammers, which are maintained by Internet service providers and public organizations (DNS-based Blackhole Lists). System administrators can add the addresses of trusted correspondents to a safe list, ensuring that their messages are always delivered without undergoing filtration.
2. **SPF and SURBL.** The filtration process also involves verifying senders using the Sender Policy Framework. Detection of spammer IP addresses using DNSBL is supplemented by SURBL technology (Spam URL Real-time Block List), which can identify spam URLs in the message body.
3. **Formal attributes.** The program recognizes spam by such typical characteristics as distorted sender addresses or the absence of the sender's IP address in DNS, an excessive number of intended recipients or hidden addresses. The size and format of messages are also taken into consideration.
4. **Content of messages (linguistic heuristics).** The program scans messages for words and phrases that are typical of spam messages. Both the content of the message itself and any attachments are analyzed.
5. **Signatures (templates of spam).** Lexical signature databases are updated round-the-clock. Using spam signatures, the program can even recognize modified versions of spam messages that have been altered to evade spam filters.
6. **Graphic signatures.** A database of signatures for graphic spam equips the program to block spam messages containing images, a type of spam that has become increasingly common in recent years.
7. **Real-time UDS requests.** The Urgent Detection System is updated with information on spam messages literally seconds after they first appear on the Internet. Messages that could not be assigned a definitive status (e.g., spam, not spam) can be scanned using UDS.

Content filtration is a particularly efficient method of identifying spam, since it allows the program to classify each message automatically.

On the basis of this analysis, the filtration server should be able to place each message into one of four categories: spam, offensive content, possible spam or not spam.

## Filtration quality and false positives

The main indicator of the quality of a spam filter is not a high detection rate of spam, but the absence of false positives (legitimate letters mistakenly labeled as spam). Antispam programs achieve a high degree of accuracy, but can never be 100% correct all the time. For this reason, it is not advisable for the system administrator to ever delete incoming messages that have been filtered using content analysis. This kind of email should be archived (for example, redirected to an address where it can be stored temporarily).

At present, Kaspersky Anti-Spam 3.0 correctly detects and removes 95 to 98% of spam messages with a false positive rate of 0.001% (1 letter in 100,000). Increasing the detection rate beyond the 98% mark is not suitable, because this would inevitably lead to an unacceptable number of false positives.

It is worth pointing out that false positives are not normally generated by business correspondence, but by news updates and messages that lean towards marketing lexicon.

Safe lists are a great aid in reducing the risk of false positives, since they allow the administrator to exclude the whole company address book from filtration, including all employees, business partners, press contacts, etc.

## Filtration database

In conducting content analysis, the filtration server uses a database of linguistic data, which downloads updates from the Internet every 20 minutes with the latest signatures added by the linguistic laboratory.

The database contains the following data types:

- a hierarchical list of spam by category;
- semantic templates;
- message signatures.

The hierarchical list includes 500 headings that are typical of different categories of spam, such as “visit our site”, “over 18”, “buy Viagra”, “enlarge your”, “consolidate debt”, “be your own boss”, etc.

Each heading has its own semantic template with a collection of phrases and a weight. There are three types of phrases in the database:

- a) **Phrases that definitely indicate spam**, such as “this is not spam!”
- b) **Phrases that are likely to indicate spam**, such as “visit our site”, “Nigeria” and “unsubscribe”.
- c) **Phrases to consider in context**, which by themselves do not indicate spam but can help confirm spam when used together with other spam phrases (e.g., brand names, phrases linked to holidays abroad, etc.).

At present, the filtration database contains around 50,000 phrases.

## Supported languages

The filtration server includes linguistic support for different languages. Spamtest technology currently supports English, French, German, Russian and Spanish.

There is additional support for detecting spam in other European languages, but the accuracy of detection is slightly lower than that for the main language sets. Using the signature method and phrases added by users, the antispam filter can work with any European language.

## Policies for processing spam

Once a message has been identified as spam, there are a number of options for processing it: delete the message, inform the sender that delivery has been refused, add a note in the subject field for the recipient, archive the message, send notifications to the administrator, end user, etc. Our flexible configuration tools allow the administrator of the mail server to decide how different kinds of messages will be processed for different user groups.

The following actions can be applied to spam messages:

- Refusal. Spam messages are not accepted by the server, just as if the intended recipient's address did not exist (the sender receives a message to this effect to discourage any further attempts).
- Deletion. Spam messages are deleted and the sender receives no notification.
- Archiving. Spam messages are redirected to an archive address and not delivered to the addressee (there is the option of sending a notification to senders of such messages).
- Delivery to the addressee with a note. Spam messages are sent to the intended recipient and marked as spam (using a note in the subject field [SPAM]), which allows the message to be processed by the mail client (for example, using Microsoft Outlook rules).

The above list is not exhaustive, but contains the most common types of processing options. The administrator can configure virtually any type of processing policy using the product's highly flexible settings.

## Attachment types

The filtration server not only analyzes the message body, but also any attachments to the message. The product supports filtration of email messages in the following formats:

- Plain text (ASCII),
- HTML,
- Microsoft Word (Versions 6.0, 95/98/2000/XP),
- RTF, and
- GIF, JPEG and PNG.