Kaspersky Anti-Spam SDK 4
Spam Detection Technologies

Whitepaper

- Basic characteristics of anti-spam solution
- Kaspersky Lab spam filtering methods
- Achieving ultimate email filtering rate
- Advantages of Kaspersky Anti-Spam technology
Copyright Information

Kaspersky® is a registered trademark of Kaspersky Lab. Other trademarks found in this publication have been used for identification purposes only and may be the trademarks of their respective companies.

This document is provided for informational purposes only and Kaspersky Lab makes no warranties, either expressed or implied, with respect to this document. All information contained herein is provided «as is», is subject to change without notice and does not represent a commitment on the part of Kaspersky Lab.

Confidentiality Notice

The information in this document may contain confidential and/or privileged material and is intended only for authorized Kaspersky Lab partners. Any review, retransmission, dissemination or other use of, or taking of any action in reliance upon, this information by persons or entities other than the authorized Kaspersky Lab partners is prohibited. No part of this document may be transferred to other parties in any form by any means without the prior written consent of Kaspersky Lab. Disclosure of any information contained herein to any unauthorized third parties is expressly prohibited.

© 2009 Kaspersky Lab. All Rights Reserved.
1 Introduction

The Spam Problem

Spam is a real problem that keeps growing each year at a furious pace. According to numerous industry reports, the volume of spam worldwide currently accounts for at least 80% of all mail traffic, increasing with each passing year. In other words, only 1/5 of all mail traffic is non-spam – i.e., business and private messages – while the rest is pure junk mail.

According to Aberdeen Group, the amount of spam in 2003 was 25% of all received email in office networks. This means that the rate of spam growth over the last five years is over 3 times the 2003 numbers. Consumers accessing the Internet are shocked by skyrocketing spam levels. Network bottlenecks and additional storage costs have led to mandates for businesses to implement effective spam-filtering.

However, network bottlenecks and additional storage costs are hardly the most harmful effects of spam. Declining employee productivity is the overwhelming by-product of spam. The resulting worldwide drag on productivity results in huge losses negatively affecting the world economy. In 2007, Nucleus Research Inc reported productivity losses from spam related incidents exceeded $712 per employee annually in the U.S. alone, totaling over $71 billion of lost worker productivity for business owners.

What is Spam?

In order to understand anti-spam technologies and decide how to effectively fight spam, we should understand exactly what spam is.

Spam is electronic correspondence with three main attributes: it is unsolicited, it is massive and it is anonymous. It is important that these three factors are included in the definition. For example, a mass mailing that a user subscribed to in the past (and that he may opt to unsubscribe from), is not spam. Mass mailings that are unsolicited but sent manually from a known address (for example, a sales representative) are not spam either.

When evaluating the spam filters used in different anti-spam products it is very important to remember that not every unwanted email is actually spam.
2 Basic characteristics of anti-spam solutions

Each anti-spam solution uses its own set of different spam detection algorithms; as a result, various anti-spam solutions have different performance levels. However, all anti-spam products have two main characteristics in common that determine their overall efficiency:

- How many spam messages the solution detects
- How many legitimate messages it marks as spam.

The first criterion is called Detection Rate, and the second is known as the False Positive Rate. Each criterion is further described below in this section.

What is Detection Rate?

The detection rate (or the percentage of spam stopped by the spam filter) is the only objective way to determine the filter’s effectiveness. Detection rate is the number of messages defined as spam by the spam filter compared to the number of all spam messages in email traffic over a fixed period of time. In other words, if your personal inbox receives 100 spam messages in 24 hours, and 5 make it past the spam filter, then the anti-spam detection rate is 95%. If a company receives 100,000 spam messages over the course of a week, and 97,000 are blocked by the anti-spam product, the detection rate is 97%.

It is impossible to automatically calculate the detection rate, since you need to know the number of all spam messages received on your mail server (which can only be totaled manually). This is why determining the detection rate can be a rather difficult, but important task.

These days, a spam filter is considered effective if it has a detection rate of 95% or higher\(^1\). However, the detection rate is not the only criterion for evaluating anti-spam solutions. Users should also consider the false positive rate.

What is False Positive Rate?

A false positive occurs when an e-mail message is erroneously tagged as spam. This presents a problem: a significant false positive rate means that normal e-mail correspondence is undermined, since legitimate messages are marked as spam and do not reach their intended recipients.

False positive rate is the number of false positives divided by the number of all non-spam messages in email traffic over a certain period of time. Like the detection rate, the false positive rate must be calculated manually.

False positives are tricky – just when you think you have done all you can to rule them out, they have a way of sneaking up on you anyway. False positives depend greatly on the quality of the spam filter. If the product demonstrates a strong detection rate, but is frequently tagging legitimate e-mail as spam, the negatives outweigh the positives and often result in big trouble for any company that values its e-mail. A few extra spam messages a day is a small price to pay, but with e-mail being a mission critical asset nowadays, the inability to receive an e-mail from your business partner or client can result in serious consequences.

Some anti-spam software vendors assert that their false positive rate is zero (which is evidently impossible) and then, in fine print, specify that it is zero only when a spam folder has been designated. A spam folder is where all detected spam is saved – it’s a critical resource for the recovery of false positives. However, if a user has to check this folder daily for false positives, it means that the spam filter is not doing its job and that the user must do most of this work manually.

The industry standard for false positive rate is 0.001% (one false positive per every 100,000 messages) or less\(^2\). A genuinely effective anti-spam solution must demonstrate such performance in live conditions, when assaulted by real world spam, not just against a handpicked, easily detected training spam set that makes the product look good only on a marketing leaflet.

---

\(^2\) See footnote on page 3
3 Kaspersky Lab Spam Filtering Methods

Spam filtering can rely either on formal methods or on content analysis based on artificial intelligence.

A formal approach uses lists of known spammer e-mail and IP addresses along with lists of open mail relays used by spammers and formal message header processing rules. The former method is used to reject letters sent from known open relays or from spammer-tolerant ISPs. The latter method recognizes messages with typical properties suggesting spam such as no recipient, too many recipients, unknown sender, etc.

Artificial intelligence (AI) techniques analyze message contents – i.e., message subject, body and attachments. To recognize suspicious content, word statistics and collocations, message content fingerprints and other methods are employed.

AI content analysis methods fit into three main categories: linguistic databases, signatures and graphical content analysis.

- **Linguistic databases** comprise keywords or terms (words and phrases) usually found in spam messages, along with their statistical data. These terms can then be allocated either into a category from a predefined set (Porno, Travel, Seminars, Diplomas, Viagra etc.) or into just one of the two categories: Spam/Not spam.

- **Signatures** are known spam sample fingerprints. These samples are gathered manually by anti-spam specialists or obtained via online anti-spam service providers from their user communities.

- **Graphical content analysis** is intended for processing of messages with attached pictures. It allows catching spam that doesn’t contain any words in message body, but tries to deliver the information only in the form of images.
Kaspersky Anti-Spam SDK uses the following basic methods to ensure recognition of all types of spam:

1. **Content analysis (Artificial Intelligence) methods:**
   - **Signatures** – linguistic fingerprints of known spam messages based on “fuzzy” comparison algorithm.
   - **Heuristics** – linguistic heuristics based on special term databases and “fuzzy” mathematics.
   - **Graphical content analysis** – image recognition technology catches spam camouflaged as an image.

2. **Formal (rule-driven) methods:**
   - **Rules** - set of formal rules based on the analysis of message header, size, sender, etc.
   - **Real-time Blacklists** – usage of so-called “blacklists” that check the message sender IP and e-mail addresses against several conventional real-time blacklists located on the Internet.

3. **Real-time spam detection technologies based on Urgent Detection System (UDS)**

   **Content Methods**
   A spam message may not have any formal spam attributes – it can be directed at a particular recipient (as opposed to undisclosed or hidden recipients) from an address that is not included in any blacklist. Content analysis algorithms are used to recognize and process such messages. The message content is analyzed using AI methods.

   Attached files in the following formats are also processed:
   - Text: plain text (ASCII)
   - HTML
   - Microsoft Word
   - RTF
As mentioned above, three Artificial Intelligence methods are used in Kaspersky Anti-Spam SDK to detect messages with "dubious" content: signature analysis, linguistic heuristics and graphical content analysis. Let’s have a deeper look into each one of them.

**Signature Analysis**

Signature analysis is based on matching of email messages against patterns from database and subsequent ranking of the messages depending on matching results. Kaspersky Anti-Spam database contains words and regular expressions inadmissible for legitimate email messages. Using spam signatures, Kaspersky Anti-Spam SDK can even recognize modified versions of spam messages that have been altered to evade spam filters.

Kaspersky anti-spam lexical signature databases are maintained by a team of professional linguists and updated every 5 minutes.

**Linguistic Heuristics**

The Linguistic Heuristics Module is used to analyze the meaning of text. It scans messages for words and phrases that are typical of spam messages. Both the content of the message itself and any attachments are analyzed. Compared to a simple keyword search, content analysis does not just look for individual keywords, but rather treats the whole text as a single entity. Based on semantic and statistical methods, each text is analyzed and compared with a database of 22 predefined categories. The category list includes such categories as Porno, Nigerian scam, Seminars, Travel, Phones, Personal Finance, Gambling among others. Additional categories are constantly added, as required.

The analysis algorithm takes into account the number, distribution and proximity values of the terms found in the message. A message can be referred to one or several database categories, along with a proximity estimate according to analysis results. When setting up filtering rules, it is necessary to define the rules for the processing of e-mail messages relating to various spam categories.

**Graphical Content Analysis**

Kaspersky Anti-Spam SDK has a dedicated technology for graphical spam analysis. When a message with graphical content gets processed, it is investigated for “suspicious” words using Kaspersky image processing technology. Thanks to advanced image processing
algorithms, this technology does not overtax computing resources as opposed to more commonly used Optical Character Recognition (OCR) technologies.

A database of graphic spam signatures enables the program to block messages containing spam images, which have become increasingly pervasive in recent years.

**Formal Methods**

**Analysis of Formal Attributes**

Kaspersky Anti-Spam SDK recognizes spam by such typical features as distorted sender addresses or lack of sender IP address in DNS, excessive number of intended recipients, or hidden addresses. The size and format of messages are also taken into consideration.

**List-Based Filtering**

This type of filtering includes checking of sender IP addresses against DNS Black Lists (DNSBLs), verification of URLs of spammer web sites in message text against Spam URL Real-time Block List (SURBLs) and verification of senders using the Sender Policy Framework (SPF). Blacklists are maintained by ISPs and public organizations, and their proper application may guarantee a reasonable spam detection rate, especially when combined with message contents scanning.

System administrators can add addresses of trusted correspondents to a safe list ensuring that their messages bypass the filter and are delivered straight to the stated recipient.

The filtering process also includes the verification of senders using the Sender Policy Framework which uses TXT records in DNS to check whether the sender domain is falsified.

**Urgent Detection System**

The Urgent Detection System (UDS) is an original “in-the-cloud” spam detection technology developed and supported by Kaspersky Lab. It is based on the following principles:
1. A message is analyzed and yields a set of properties that can be used to identify it. These properties may include header information, text fragments and other information about the message being processed.

2. The filtering server uses the properties collected during analysis to generate a small UDS request and sends it to one of Kaspersky Lab’s UDS servers. This request includes only a hash of the message which precludes a third party from viewing the recipient’s processed mail. The collection and delivery of this information poses no risk to data confidentiality or privacy.

3. The UDS server checks the received request against a database of known spam. If the request matches a known spam sample, a message is sent to the filtering server informing that the e-mail is very likely to be spam. This information is then taken into account during the assignment of a specific status to the e-mail.

Figure 1. Urgent Detection System within Anti-Spam workflow

Thanks to 24x7x365 human analysis of spam from Kaspersky Lab, UDS is updated with spam message signatures literally seconds after they first appear on the Internet. As a result, this technology allows real-time filtering of known spam, before regular updates (every 5 minutes) to the anti-spam databases become available.
4 Achieving Ultimate Email Filtering Rate

As addressed previously, the effectiveness of each spam recognition method can be calculated using two primary criteria, detection rate and false positive alarm rate which is defined as the precision value. Some anti-spam vendors largely rely on real-time methods of spam detection provided by “in-the-cloud” services. Content methods aren’t a strong suit of such solutions, which will result in a zero detection rate in case the “in-the-cloud” infrastructure becomes inaccessible. A combination of formal, linguistic, and real-time methods – the approach chosen for Kaspersky Anti-Spam SDK – is the best way to obtain a competitive detection rate and nearly eliminate harmful false positives.

Fine-tuning anti-spam systems on-site may raise the detection rate to 97-98%, although considerable care must be taken here. A spam filter’s task is to decrease the flow of unwanted mail. Nevertheless, 100% detection of unwanted messages cannot be guaranteed because excessively strict criteria inevitably result in the filtering out of useful information.

In our expert opinion, the most important criterion for anti-spam tools is not the highest detection rate, but rather the lowest possible probability of false alarms combined with a reasonable detection rate. For this reason we do not consider the spam detection rates of 99.5% or even 99.8% (as stated by some anti-spam vendors) as meaningful. An anti-spam engine that manages a 95+\% detection rate while not exceeding the 0.001\% false positive probability is of much greater overall value.
5 Major Advantages of Kaspersky Anti-Spam Technology

The primary advantages of Kaspersky anti-spam detection technologies include:

- All detection methods are used simultaneously, yielding high detection rates and remarkably low levels of false positives
- An up-to-date, comprehensive database of linguistic heuristics enables intellectual content analysis and thus the recognition of brand-new spam samples including polymorphic spam
- A top-notch reaction time achieved by combining content analysis with anti-virus-like regular updates schemes
- All employed detection methods are controlled from a single administration point that enables on-site fine-tuning and customization

24x7 Anti-Spam Laboratory

Kaspersky anti-spam technology is supported by human analysis. The team of professional linguists works 24x7x365 to analyze the global spam weather and to continually develop new spam filtering methods and rules. Why is this important? To the best of our knowledge, only expert human analysis makes it possible to achieve minimal rates of false positives combined with rapid reaction to new spam techniques.

Unlike many other anti-spam solutions, the anti-spam engine used in Kaspersky Anti-Spam SDK requires minimal user interaction and no training during setup and operation. The Kaspersky Anti-Spam Laboratory processes enormous volumes of spam, collected from all over the world, constantly adding new spam samples to signature databases and analyzing new spammer tricks. This combination of high-class expertise, human-created signatures and rules enable the solution to detect wide-spread spam attacks. Detection remains reliable even if the message is slightly changed (as in polymorphic graphical spam), since manually-created signatures allow covering all varieties of messages sent from a single distribution point. Human analysis with an enhanced anti-spam engine together provides excellent detection rates and accuracy.
6 Contact Details

We welcome all Anti-Spam SDK partnership inquiries. As the Anti-Malware SDK industry leader, we offer extremely flexible and competitive business practices ensuring the best possible partnership experience. For further information or engagement with our global Business Development staff, please contact us at your convenience. Our contact information is listed below.

<table>
<thead>
<tr>
<th>Mailing Address:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10/1 1st Volokolamsky Proezd</td>
<td>123060 Moscow</td>
</tr>
<tr>
<td></td>
<td>Russian Federation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.kaspersky.com/oem">http://www.kaspersky.com/oem</a></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OEM and Technology Licensing:</td>
<td><a href="mailto:oem@kaspersky.com">oem@kaspersky.com</a></td>
</tr>
<tr>
<td>General Information:</td>
<td><a href="mailto:info@kaspersky.com">info@kaspersky.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tel.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+7 495 797 8700</td>
<td></td>
</tr>
<tr>
<td>Fax:</td>
<td></td>
</tr>
<tr>
<td>+7 495 645 7939 or +7 495 956 7000</td>
<td></td>
</tr>
</tbody>
</table>