ARSKA - Arrak Spam Killing Appliance

White Paper

© Arrak Software, March 2014

Oy Arrak Software Ab
Tillinmäentie 1 C
02330 Espoo
FINLAND

tel +358 (0)9 855 0365

www.arrak.fi
info@arrak.fi

Table of Contents

1   Introduction

1.1   Purpose

The purpose of Arska is to provide a black-box approach to filtering of unwanted emails. The filtering consists of stopping unwanted spam as well as messages containing malicious code or viruses before they enter the mail server of the company. The focus is on email arriving from external sites going into the organization, but outgoing email can also be filtered.

The black-box mail gateway works on the SMTP protocol level, and is therefore interoperable and compatible with any mail server.

1.2   Architecture

A rough picture of the solution would be as follows:
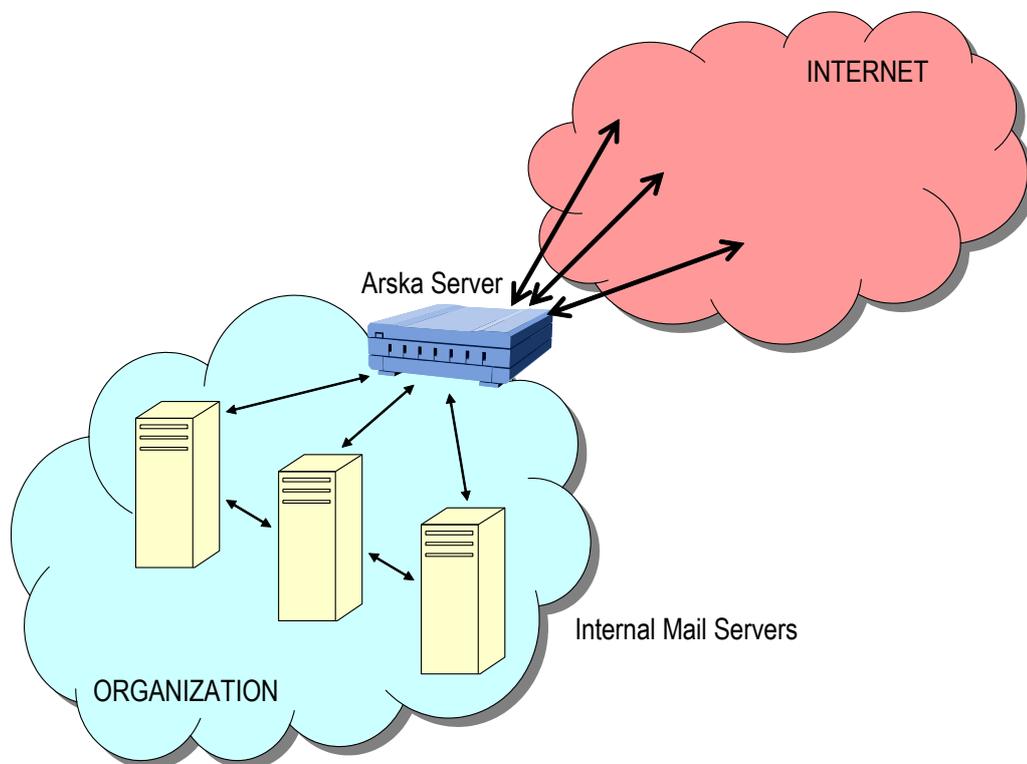
Figure 1. Architecture

The Arska server sits on the border to the organization in such a way that all incoming email must pass through it. If the email is accepted, it is forwarded to one of the internal mail servers for delivery to the final recipient. Mail sent within the organization does not pass the gateway, whereas outgoing emails must again pass the gateway (to guard against sending virus-infected emails to customers). The outgoing filtering also protects against cases when a virus-infected or hacked computer within the organization is used by spammers as a sending host for more spam.

2 Technical Solution

2.1 Components

Arska consists of a stand-alone Linux-based server containing reliable third-party software that Arrak Software has many years worth of experience using, such as:

- **Exim** mail software (http://www.exim.org)

- **Amavis** mail scanning software (http://www.ijs.si/software/amavisd/)

- **SpamAssassin** spam checker (http://www.spamassassin.org)

- **Gross** Greylisting of Suspicious Sources (http://code.google.com/p/gross/)

- **Pyzor** (http://pyzor.sourceforge.net)

- **ClamAV** anti-virus (http://www.clamav.net)

- **PostgreSQL** (http://www.postgresql.org)

- **Apache** HTTP Server (http://httpd.apache.org)

- **Mason** (http://www.masonhq.com)

- **OpenSSH** (http://www.openssh.org)

- **Sender Policy Framework** (http://spf.pobox.com), **DKIM** DomainKeys Identified Mail (http://www.dkim.org) and Microsoft Caller-ID

- as well as additional components, integrations and functionality by Arrak Software

All listed third-party software is freely available, and represent state of the art software with regard to security, performance, stability and functionality. As new techniques emerge that can help fight spam or improve security, Arska can easily be upgraded to include the newest technologies when they have been sufficiently validated and tested.

2.2 Virus Protection

Arska has the possibility to use multiple antivirus softwares in addition to the **included open-source ClamAV** software. The use of multiple anti-virus softwares is becoming increasingly important, as outbreaks of new viruses can spread rapidly during the critical hours before anti-virus software are updated to recognize the new virus.

Note that the use of collaborating spam tracking databases, such as DCC and Pyzor, can protect you against new aggressive viruses long before normal anti-virus software is updated to recognize viruses. This is because new viruses that spread quickly via email are stopped when e.g. the collaborating databases detect the same email being sent to a large number of users all over the world.

Other spam tests are also quite effective for stopping new virus outbreaks. Virus-generated messages share a number of properties also seen in spam, such as fake sender addresses, malformed email headers, disguise attempts, bad SMTP implementations and other subtle details that all can add up to a sure indication of non-wanted email.

As a sensible precaution, Arska can block file attachments with undesirable extensions. There is generally no need to allow emails with .wav, .scr or .pif attachments, or attachments with double extensions or names trying to lure the user into opening some seemingly innocent file.

2.2.1 Retroactive Virus Scanning

Despite the above precautions of trying to minimize the potential risk of getting virus-infected emails, there is still a chance that new viruses are passed through before any virus-scanning software gets updated to recognize the new virus. When new antivirus-databases have been downloaded, Arska has the ability to perform a retroactive scan from an internal cache of previously delivered emails, and send warning messages to recipients that have received infected emails prior to the update. In addition to providing post-incident warnings, this also gives exact data about the efficiency of antivirus software updates.

2.3 Security

2.3.1 Transport Encryption

Arska comes with its own strong TLS (SSL) certificate, meaning emails are both received and transmitted using **encrypted connections** whenever possible (requires industry standard TLS support by the remote SMTP server). It is also possible to configure domains for which TLS encryption is required, **building a network of secure domains** for which secure email delivery is enforced. Emails going to a secure domain can only be delivered to the correct server, prohibiting sensitive emails being eavesdropped, routed via untrusted servers or delivered to the wrong server.

2.3.2 Sender Authentication

Arska can optionally be configured to provide DomainKeys Identified Mail (DKIM) **signing of outgoing emails**. DKIM is completely transparent to the end user and provides message header and body integrity verification using strong cryptography to the remote mail server. I.e. the receiving server can verify that the DKIM signed email is intact and has been sent by the domain that has signed the email.

Likewise Arska can **validate incoming DKIM signed emails** and use the reputation of a domain for additional spam scoring. Trusted domains can be whitelisted, while emails without a valid DKIM signature that try to look like they come from a domain that normally signs their emails can be given spam points.

2.3.3 Attack Protection

Arska is protected against brute force or dictionary attacks (attempting to guess email addresses by trying), unauthorized relaying and sender spoofing. Any such attempts will result in automatically slowing down and denying access from the attacking host. Overly abusive IP addresses can also be automatically firewalled for periods of time, which will further protect Arska against attacks.

2.4 Performance

The additional time required for an email to pass the Arska-server is typically below 10 seconds (most of the time is spent waiting for results from network tests such as DNS-based blocklists), so the delay introduced by the extra mail gateway is hardly noticeable.

A single Arska-server should be able to handle volumes of up to at least 150 000 messages or 5 GB of data per day. Performance and filtering accuracy are opposing principles - you cannot achieve great accuracy without a performance hit. Arska places a greater significance on the filtering accuracy, resulting in a relative low performance for a mail gateway. However, the performance of a single Arska server should be sufficient for an organization of up to some thousand users (or a bit more or less depending on email habits and spam amounts). A gateway capable of processing 10 million messages a day cannot be as accurate as a gateway that does a lot more processing per message to classify messages.

If redundancy or more performance is needed, **multiple Arska-servers can be set up** that share a single management interface with aggregate views of the whole cluster.

2.5   Confidentiality

End-users have the possibility to access a web-interface in Arska, showing how messages sent to them have been processed. There is no need for any administrator to access the emails of other users, as the end-users themselves have full access to their quarantined messages and can release any or their stopped messages if they so wish.

Administrative logins can also be created, granting access to aggregate reports and performance data for whole domains, without revealing individual messages.

2.6   Reliability

There are many aspects of reliability that have been considered in Arska:

- **hardware** - thoroughly tested hardware with RAID-mirrored disks (or running as a virtual server on a reliable host using VMWare or Hyper-V).

- **software components** - regular internal self checks that all components are working properly and receive updates on a timely basis (such as anti-virus databases). Alerts will be raised if some component is not working properly or if data is not being updated.

- **mail processing priority** - the normal flow of messages should not be delayed or hindered if there are problems with network-based tests, nor is the SQL database involved during mail processing (to avoid potential bottleneck problems). Actions in the web-based interface cannot interfere with normal mail processing.

- **mail responsibility** - ensure that once a message have been received, Arska is responsible for the delivery or quarantine of the message. It cannot just disappear without a trace, even under fatal conditions (such as power failures)

- **quarantine notifications** - Arska can send digests to end-users about messages that have been placed in the quarantine, so users can be aware of falsely stopped messages. Senders also get a non-delivery report for messages that have been stopped (unless the spam score is above a certain limit)

- **filtering sensitivity and specificity** - it is more important to recognize non-spam than to stop everything containing offensive words. A legitimate message being falsely stopped is far more expensive than a spam slipping through the filter, so spam-filtering can not be too aggressive. Unsure cases are delivered to the recipient with a clear SPAM-mark in the subject.

- **disk space management** – all quarantined emails are stored as long as there is disk space available. Oldest quarantined emails are automatically deleted once the disk space is low enough, but as most spam is rather small (and compressed even further) a typical Arska server can hold quarantined emails for at least a year (typical customers have space for 3-5 years).

2.7   Installation

2.7.1   Hardware

The Arska server is installed in the organization's network at a suitable logical location. This can be either outside the firewall, inside the firewall or in a separate DMZ network, depending on what is most convenient for the network administrators of the organization. The Arska server is itself secured by a built-in firewall, so it is perfectly safe to have Arska on the outside of the organization's firewall directly accessible from the internet.

2.7.2   Mail Routing

The mail routing is changed so all incoming mail is directed to the Arska server, i.e. the primary MX records for the domain should point to the Arska server. The firewall in the organization

should be set up in such a way that no mail can be delivered directly to the internal mail servers from outside the organization, but that mail can be delivered over SMTP to the Arska server, which in turn can forward the filtered mails to the internal mail servers using SMTP. Arska can route emails to different internal servers based on the recipient domain or using randomized or priority based alternatives.

It is also recommended that all outgoing email is routed through the Arska server, as this protects against infected or hacked workstations sending viruses or acting as spam senders.

2.8 Customization

One of the largest benefits of a **customer-specific server** is the ability to customize the filtering according to both customer needs and their circumstances. Accuracy can be improved by **tailoring the filtering rules** to explicitly allow what is considered non-spam and filtering out the particular non-wanted stuff.

A particular example are the customized rules we always write for each customer to **classify bounce messages** based on if they are bounces for something really sent by the organization, or if it is a false bounce. False bounces (also called collateral spam or backscatter) is the result of spam or viruses being sent with a false sender addresses, causing error messages and virus warnings to bounce to innocent users, which is a nuisance and causes unnecessary concerns ("Hey, some system is telling me I have sent them a virus and they recommend I disinfect my machine!"). Our customized rules look at small hints in the bounces to correctly identify if the bounce is valid, so you actually get the error messages for messages you actually have sent while false bounces can be safely filtered out.

2.9 Add-ons

Because Arska is built on a standard Linux platform, there is an almost infinite range of additional software that can be installed to satisfy customer needs. Some examples:

- a web-based interface ("webmail") for accessing the internal mail server via e.g. IMAP

- LDAP-connectivity for obtaining valid recipient addresses from an internal server, so Arska can immediately reject invalid recipients

- public DNS server with full management via web interface and new DNS zones can be automatically replicated to other Arska servers

- mailing list manager with web-based interface for administering lists. List members can also access web interface to manage own subscriptions and browse mailing list archives

- SNMP monitoring of Arska operational status

- integration of custom reports to organizational data systems, i.e. the internal SQL database containing all message-related statistics can be exposed to internal reporting systems for full insight into all data gathered by Arska. A cluster of Arska servers share a single SQL database, giving a centralized view of the total mail flow

3 Operation

3.1 Mail Processing

The Arska server processes all emails in the following way:

- Incoming SMTP connections are checked that they adhere to normal mail standards. Attempts to overload the server by brute-force guessing of mail addresses, giving massive amounts of syntactically invalid commands, or trying

to relay messages to other organizations is automatically stopped. Overly abusive senders can be firewalled for a period of time.

- SMTP connections from servers listed on several ip-based blocklists are greylisted on the first connection attempt, i.e. given a temporary failure and required to retry delivery after a while. Most spammers do not retry (although it is a requirement of proper email servers to be able to do so), and therefore 70-80% of incoming spam can be avoided without further processing.

- Accepted SMTP connections result in the email being received by the server for further analysis. Arska does not acknowledge receipt of the email until it has been persistently stored on disk, i.e. responsibility of the message is not assumed until it can survive any kind of server disaster.

- The email and all its attachments are extracted, and all parts are scanned by the anti-virus software for dangerous code. If the email is determined to contain something dangerous, the email is archived in a quarantine area and a notification is sent to an administrator and to the original sender (unless the email is a virus that is known to use fake sender addresses).

- The email is analyzed in a vast number of different ways (see section 3.2), to determine whether it is a genuine message or some form of unsolicited or bulk email (spam). The analysis results in the email being assigned a numerical spam score.

- If the spam score exceeds a kill limit (typically 8-10), the email is archived in a quarantine area and the original sender is sent a delivery failure message (only if the sender address indeed can be inferred to be correct using SPF or DKIM as most spam use spoofed sender addresses and we should not send any error messages to those innocent victims)

- If the spam score exceeds a slightly lower limit (typically 5), the email is passed on, but with a clear indication in the subject, as well as in special headers, that it is spam.

- If the spam score is above a lower limit (above zero), some special headers are inserted in the email to alert the recipient that the mail might possibly be spam, and the mail is passed on.

- Email with a low enough spam score (negative score) is simply passed through, with only a header line added that indicates that the email has passed the spam and virus checking.

- Passing on an email means the email is delivered over SMTP to the internal mail server based on the recipient domain. For each recipient domain one or several internal mail servers can be configured in priority- or randomized order.

So for every email received, **either of two things happens:**

1) The **email is delivered** to the internal mail server, but there might be some extra headers or markings in the subject to indicate that it might be spam.

2) The **email is not delivered** to the internal mail server, in which case the message is stored in a quarantine area in the Arska server (can be accessed later), and the original sender (if trusted) gets a **notification** why the delivery has been stopped.

Incoming emails never disappear without a trace!

3.2 Spam Analysis

Performing spam analysis on an email includes **a vast variety of different tests**. Each test can result in a positive or negative score being added to the total spam score of the email. This way, a perfectly legitimate mail can mention genitalia or drugs and still not be classified as spam, because there are other tests that recognize the validity of the email through the proper email headers and the context in which the bad words are used.

The different spam tests applied to each email include (but are not necessarily limited to):

- Mail header analysis: detects forged senders, open relays, blacklisted mailers, etc.

- Body text analysis: typical phrases, unsubscribe-links, shouting, images, foreign languages, obfuscation attempts, etc.

- DNS blacklists: real-time blacklisting of open relays, dialup-users, spamvertized domains, etc.

- Sender whitelists: known trusted senders

- Bayesian token classification: automatically classifies good and bad words from seen messages and their previous classification as spam or non-spam

- CRM114 phrase classification: a more advanced version of statistical Bayesian content classification that can learn to separate between good and bad phrases (up to five consecutive words).

- Collaborative Spam Tracking: check with world-wide services if the same message has been reported as spam by other users

By using a series of different tests that all contribute to the total spam score of the email, the total reliability is greatly improved, as **no single test can by itself dominate the result**. The amount of points a single test can affect the total score is **weighted according to the reliability** of the individual test. The spam tests mentioned above can be divided into three different types of tests:

- **Built-in tests** that can be performed without any external access, such as recognizing different kind of text obfuscations, spam phrases, known scams, etc.

- **Adaptive tests** that over time learns the difference between valid emails and text contained in spam – this is further described in section 3.2.1

- **Network tests** that in real-time query global servers – some of these are DNS-based blacklists of open relays, dialup-users and known spam senders, and some tests are so called Collaborative Spam Tracking (see section 3.2.2)

Naïve blocking by blacklisting single offensive word or specific senders **is not performed**, because such simple filtering is known to cause unwanted **false positives**, i.e. legitimate emails being wrongly identified as spam.

3.2.1 Bayesian Classification and Training

The Bayesian classification algorithm is able to **learn from seen emails** whether they are spam or not, based on statistical analysis of the text. This learning is mostly automatic, because scanned emails that get a high enough spam score is automatically trained as being spam, while emails with a sufficiently low score is trained as not being spam. The middle ground is a bit trickier, and therefore uncertain cases are always referred to a human user.

In this case it means that emails that get a rather low spam score is let through to the recipient, but with a special mail header. It is recommended that mail clients be configured to look for this

header and highlight emails that have this header. It is then up to the user to manually classify these emails as either being spam or non-spam, either by forward the mail to either of two addresses set up to feed the corresponding type of email to the Bayesian learning process, or by marking the emails via the web-interface. After being trained with a number of sample emails, the Bayesian classifier quickly learns to distinguish these uncertain cases and then they are not uncertain anymore.

3.2.2   Collaborative Spam Tracking

Collaborative tracking services such as DCC and Pyzor work by calculating several numeric fingerprints (also called digests or hashes) from each email, and then asking global servers if these fingerprints correspond to messages reported as spam by other users. As someone using the same collaborative database reports an email as spam, all other users are automatically protected from the same message. There are currently millions of users using one or several of these collaborative spam tracking databases, resulting in a broad network of **users helping to recognize and stop spam**.

The score added to an email if a collaborative database identifies a message as spam is weighted so it is not possible for other users to completely determine the fate of your email. A message has to collect enough points from several different tests in order to be blocked.

3.2.3   Spam Scores and Thresholds

As each email undergoes a lot of tests, the result from each test can subtract or add up to 5 points to the total score. The individual score for each test have been optimized by running a genetic algorithm on a sample of about half a million of emails manually sorted into spam and non-spam, resulting in a weighting where more significant or reliable tests score higher points than less significant tests.

Based on a sample of 740226 emails received during November 2009, 654685 messages (88%) were classified as spam. The average score for non-spam messages were -11.1, while the average score for spam messages where +42.2. A frequency distribution of the spam scores can be seen in Figure 2.
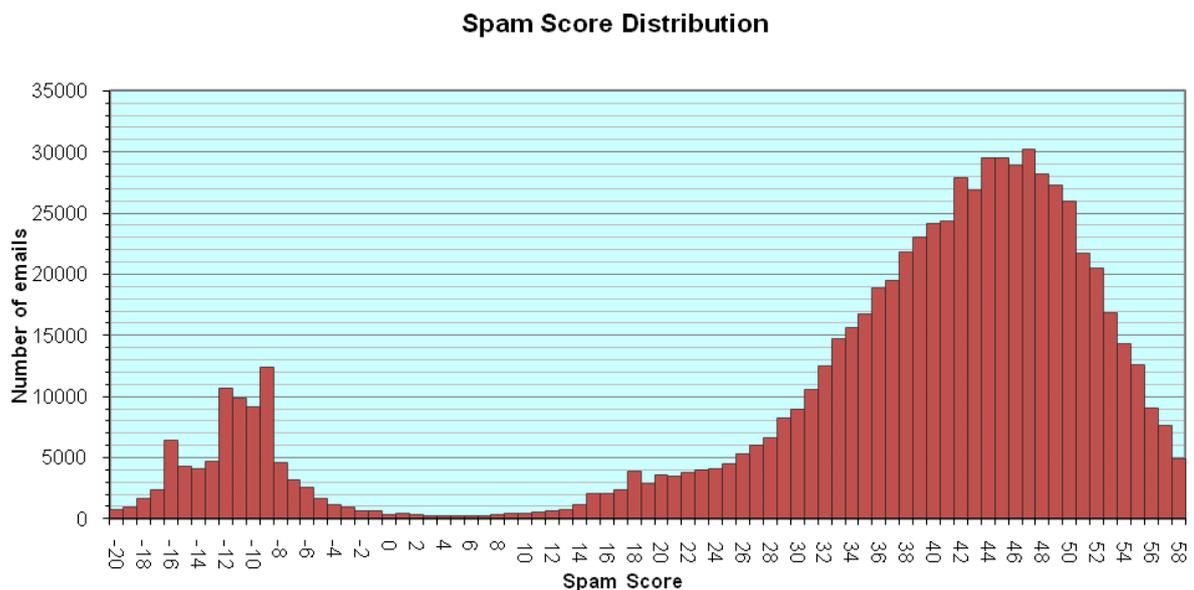
Figure 2. Spam Score Distribution from live data.
Total number of emails was 740226, of which 88% were classified as spam.

As can be seen in the figure, most of the spams get very high total scores, clearly distinguishing them from legitimate emails with negative scores. The cumulative effect of using many different spam tests contributing to the final spam score results in **something resembling two normal distributions** for the two different classes of emails (spam and legitimate emails respectively),

with a small overlap in a "grey" area between the two. This area corresponds roughly to scores in the range between 1 and 8 where there is risk of misclassifying the email. However, less than 0.3% of the total messages received fell in this unsure area.

Efficient filtering involves finding a good balance between stopping as much of the spam as possible while keeping the risk of stopping legitimate mails to a minimum. This is accomplished by careful tuning the following three cutoff levels:

- Kill level – messages with score above this are stopped (placed in the quarantine). Default value is 8.

- Tag level – messages with score above this get a clear indication in the subject (e.g. "***SPAM***") that this is probably spam. Default value is 5.

- Warning level – messages with score above this get a special mail header to indicate it is in the uncertain zone. Default value is 1.

All threshold levels are fully adjustable globally, and can be further refined by domain as well as overridden for individual users (end-users can make personal adjustments via their own account). The filtering levels can initially be set rather liberally (i.e. let through some potential spam just to make sure no legitimate emails are stopped), and then gradually adjusted to more suitable levels as the Bayesian learning improves the classification. The idea behind the tag and warning levels is to alert the user about uncertain classification – manually classifying these messages (by reporting them as spam or non-spam) improves further classification.

3.2.4   Exceptions

It is of course possible to manually add exceptions to the scanning process, such that certain friendly organizations or known sender addresses are not scanned for spam.

3.3   Measuring the Results

There is no single number that can measure how good a spam filter is. Someone claiming 99.9999% accuracy is not telling the whole story, because if they say that only one message out of a million has been stopped by mistake, they do not tell you how many messages they failed to stop (nor do you know anything about undetected mistakes). The performance of a spam filter can be more accurately described with the following four measures:

- **Positive Predictive Value**: If the filter says it is spam, how likely is it to actually be spam?

- **Negative Predictive Value**: If the filter says it is not spam, how likely is it to actually not be spam?

- **Sensitivity**: If it's actually spam, how likely is the filter to say it is spam?

- **Specificity**: If it's actually ham, how likely is the filter to say it is ham?

An overall single aggregate measure that omits some details would then be:

- **Efficiency**: The percentage of mail that was correctly classified.

The reason most spam-filter vendors do not give you these numbers, are because either they do not have the data, or it would make them look bad. All those measures require active feedback from the end-users, because without the information about what messages have been misclassified by the filter, you cannot calculate any of them!

End-users can easily tell Arska what messages have been misclassified and in what way. In addition to learning the Bayesian databases, this also gives the data required to calculate those numbers. All the listed measures are therefore available at any time.

3.4 User Monitoring

Emails that have been quarantined in the Arska server are accessible for manual processing using a web interface. Should a user suspect some legitimate email has been stopped (because the organization had a too strict kill level set), he can access the quarantine area to check stopped messages and manually release misclassified emails (which will then also be fed to the Bayesian classifier as well as revoked from the collaborative databases to improve classification of future emails).

The user can also configure quarantine summary messages to be sent to him at regular intervals or when a certain number of emails have been quarantined. The summary message lists the date, time, sender, recipient and subjects of recently quarantined emails which quickly tells the user if something has been stopped by mistake.

3.5 Server Monitoring

The performance of the Arska server is continuously monitored. Performance and information about the operation can be plotted in graphs over different time intervals (day, week, month, etc.), to quickly see the amount of emails processed, number of spam stopped, processing times, etc. More configurable reports with actual numbers and percentages can be generated with user-definable filtering and output columns. See chapter 4 for a more detailed description of web interface.

Arska also connects to Arrak Software's central server every five minutes to tell if there are any abnormal conditions (e.g. large email queues, or other changes that might require administrative intervention), and can download updated versions of common spam rules if available. Arrak Software will also automatically get alerts if any Arska server suddenly stops responding and can take proper action. Most problems can be automatically corrected by the internal self-check routines, such as freeing disk space by removing old quarantined spam, restarting software components if they for some reason stop responding, and even performing a physical reset of the server if it should hang (hardware watchdog).

3.6 Maintenance

Regular updates are an essential prerequisite for running things smoothly. Security updates and new improved software versions are important, especially with regard to the spam filtering, as is it a constant battle between the good guys and the bad guys. Spammers are continuously trying to circumvent spam filtering rules, so spam rules must regularly be updated to better combat new forms of spam (because there is more to spam filtering than just the auto-learning Bayesian classification). As part of Arrak Software's turnkey solution, we take care of all server maintenance. You just enjoy the service!

4 Web Interface

4.1 Introduction

The web interface requires users to log on with a username and password. Each user account can be granted access to different sections of the interface, but the account also determines what data can be seen in each section. More powerful users can run reports that include summary data for one or several domains, while ordinary end users only have access to data regarding messages sent to their email address. The system administrator can create any kind of account, and even delegate the responsibility of account creation to other accounts (domain administrators), which can only manage account within their realm and rights (i.e. an administrator cannot create accounts with more rights).

There is no need for an administrator to create ordinary end user accounts, as they can be created on the fly by the end users themselves. They can order a password to their email address, giving them a very limited access to only list messages sent to them, train them as

spam or non-spam, releasing quarantined messages, order quarantine notifications be sent to them via email according to certain criteria or adjust tag- and kill-levels for emails sent to them.

4.2 Interface Sections

The different sections of the Arska web interface are briefly presented below. Note that users only have access to sections according to their account rights, and the data visible in each part is limited to those email addresses or whole email domains that each account has access to. Accounts created on the fly by end users only have access to "Messages", "Training" and "My Settings".

4.2.1 Messages

Search for incoming email messages using any combination of criteria, such as received timestamp, sender, subject, score, classification, etc. In the list of matching messages you can see how Arska has classified each message (spam, virus, unsure, etc.), and you can manually mark messages as spam or non-spam (the latter also releases the message if it was quarantined). Spammy messages can also be viewed directly, which also gives a more detailed report of which spam tests were triggered for that message.
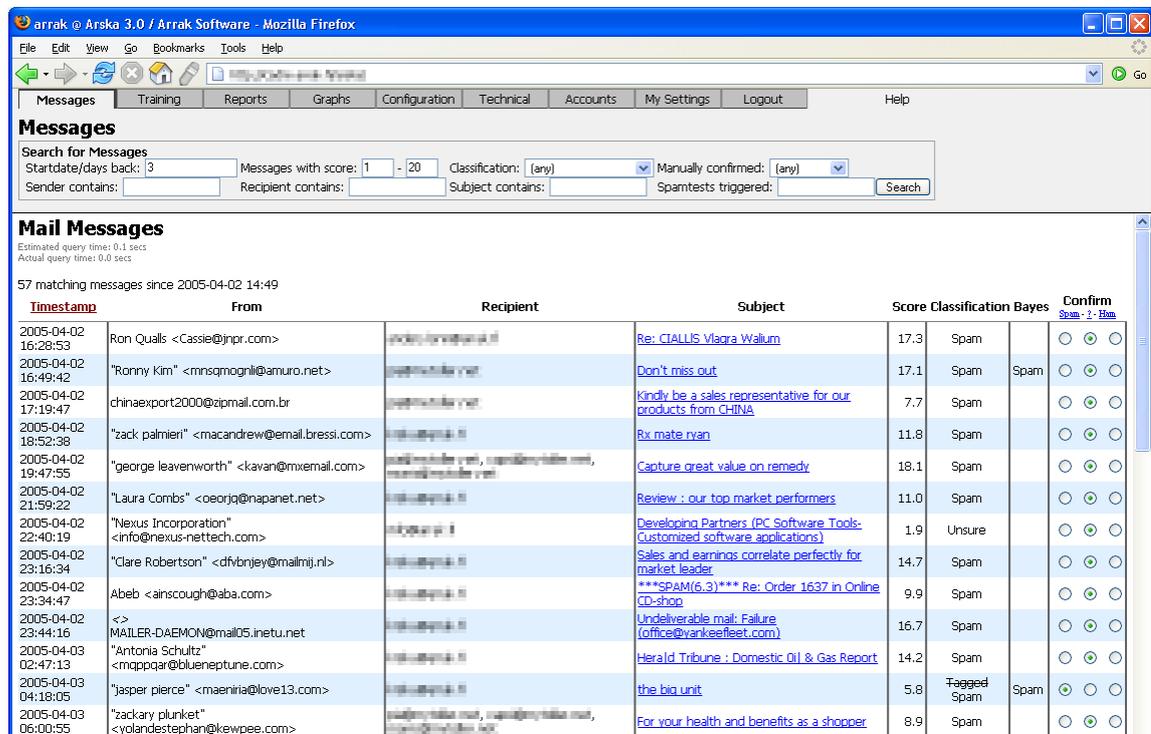
Figure 3. Messages example

4.2.2 Training

The Training section in Arska can be used to find messages that require manual classification the most, i.e. a simple priority scheme is used to find unsure messages as well as messages where the Bayesian result differs from the total spam score. The list of messages both looks and functions the same way as the list in the messages view, but this list is the quickest way to find the messages that add most information to the training process.

4.2.3 Reports

The Reports section contains different types of numeric reports that can be generated for selected time intervals, including a generic report where you can freely choose columns to be included in the report. The result can be resorted simply by clicking a column heading.

Generic Report
Number of days: 14   Messages with score: [   ] - [   ]   Classification: (any) ▾   Recipient contains: [   ]
Display columns: Date ▾   (none) ▾   (none) ▾   [Run]

Arska Report

Estimated query time: 0.0 - 3.8 secs
Actual query time: 2.0 secs

|	Date	Emails	Delivered	Deliv%	Spam	Spam%	Virus	Vir%	Rejected	Rej%
1.	2009-11-23	12917	1110	8.6%	9834	76.1%			1990	15.4%
2.	2009-11-24	33686	4580	13.6%	22806	67.7%	1	0.0%	6357	18.9%
3.	2009-11-25	39722	4879	12.3%	26755	67.4%			8145	20.5%
4.	2009-11-26	31948	4161	13.0%	22414	70.2%	5	0.0%	5410	16.9%
5.	2009-11-27	28432	3849	13.5%	21844	76.8%			2771	9.7%
6.	2009-11-28	21067	829	3.9%	18059	85.7%			2208	10.5%
7.	2009-11-29	19379	819	4.2%	15941	82.3%	2	0.0%	2639	13.6%
8.	2009-11-30	29480	4908	16.6%	20598	69.9%	1	0.0%	4010	13.6%
9.	2009-12-01	27430	4662	17.0%	18314	66.8%	559	2.0%	3962	14.4%
10.	2009-12-02	25381	4708	18.5%	17897	70.5%	5	0.0%	2832	11.2%
11.	2009-12-03	25703	3950	15.4%	19429	75.6%	7	0.0%	2366	9.2%
12.	2009-12-04	25430	3620	14.2%	19424	76.4%	5	0.0%	2438	9.6%
13.	2009-12-05	17315	750	4.3%	14794	85.4%	3	0.0%	1790	10.3%
14.	2009-12-06	17464	763	4.4%	14781	84.6%			1950	11.2%
15.	2009-12-07	17864	3028	17.0%	12521	70.1%	117	0.7%	2224	12.4%
	Total:	373218	46616	12.5%	275411	73.8%	705	0.2%	51092	13.7%

Figure 4. Arska Generic Report.

4.2.4   Graphs

In the Graphs section, many aspects of the email processing and the physical server can be plotted for different time periods (day, week, month, etc.). The graphs give a visual overview of the amount of messages received, delivered, spam ratios, processing times, server load, temperatures, etc.
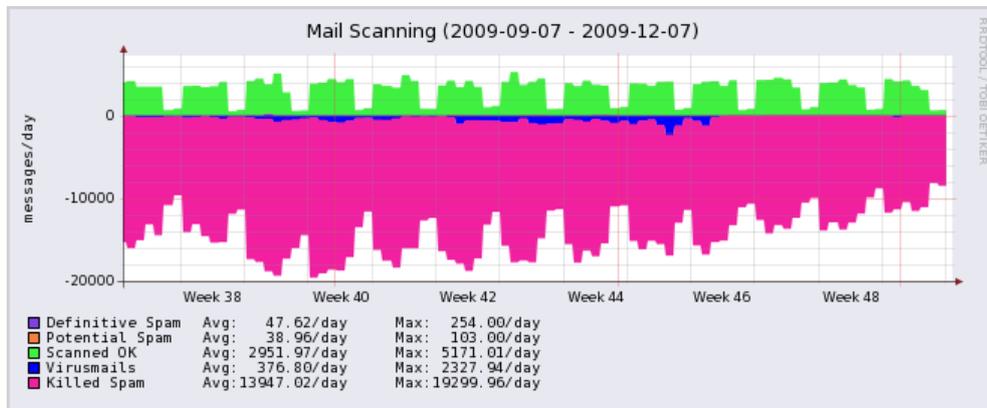

*Figure 5. Mail Scanning during the last 3 months*
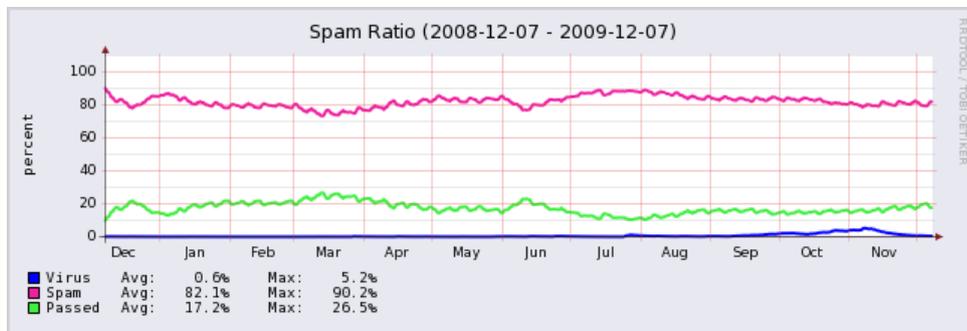
Figure 6. Spam Ratio during the last year

4.2.5 Configuration

The Configuration section shows the local domains, current filtering levels, Bayes database status, etc. This section also contains pages for explicitly rejecting recipient addresses (if Arska has no active LDAP integration) or allowing addresses (for data missing from LDAP).

4.2.6 Technical

The Technical section contains more technical reports that primarily are used by the service provider (Arrak) when tuning the filter. This section has reports for individual spam test performance, spam score distributions and filtering accuracy.

4.2.7 Accounts

User accounts for the web interface are created and edited in this section. Account properties that can be edited include username, password, quarantine notifications, associated email addresses, section access and domains that the account can administer.

4.2.8 My Settings

My Settings allow the logged on user to manage certain properties of his own account. These include changing username, password and quarantine notification parameters. Arska can send notifications to the user either when a certain number of messages have been placed in the quarantine, or after a specified time period, optionally filtered by a max score.

4.2.9 Help

Every section has its own help page, describing the purpose of each section, what can be done, and how to interpret the displayed data.