Spam Configuration/Training Guide

GWAVA4

http://www.gwava.com

TABLE OF CONTENTS

OVERVIEW

Spam is overwhelming email systems everywhere. Anti-spam solutions simply are not keeping up with the changing spam messages that are infesting mailboxes. Spam hinders the productivity of employees and annoys employees in today's electronic industry.

In 2004 Beginfinite released the product GWAVA 3, a software spam filter used on email servers. GWAVA 3 could correctly identify a good majority of spam. GWAVA 3 used heuristic rules to find and block messages that had spam content. A heuristic rule is a pattern written in regular expression that matches typical spam content in email. In addition to heuristic scanning, GWAVA 3 also provided plain text filters, attachment blocking, virus scanning, and a myriad of other features that all worked together to identify and delete unwanted messages from mail systems. GWAVA 3 remained an effective anti-spam solution until mid 2006 when new waves of spam started to hit mail systems. GWAVA 3 was around 95% effective until these new waves of spam.

Towards the end of 2006, spammers started to change their tactics to include embedded images in their email messages. These embedded images, which are nothing more than text converted to GIF images, are able to slip by GWAVA 3 virtually undetected. GWAVA 3's heuristic spam engine could not "see" the words in the embedded images. Other products have tried to use some OCR techniques to detect the new spam messages, but the scanning performance is severely crippled and OCR technology isn't very reliable. In addition, others have tried implementing new heuristic rules to stop image spam, but it raises the false positive catch rate of the spam engine. A false positive is a legitimate message that is blocked.

OCR technology might catch some image spam, but anti-spam technology would only be effective until spammers changed their tactics. The only good solution is to get ahead of the spam curve, not simply react to what spammers come up with.

In response to image spam and other new spam messages, Beginfinite has released GWAVA 4, which utilizes a new auto-adjusting spam engine that will catch image spam and any other spam messages that may come out in the future. GWAVA 4 should provide a permanent solution to the spam epidemic.

Since GWAVA 4's spam engine will auto-adjust, it is very important that GWAVA 4 is configured properly. In addition to the proper configuration, GWAVA 4's spam engine must be trained with examples of good and bad messages.

This guide will show how to properly configure and train GWAVA 4's spam engine to stop spam with a high catch rate and a low, nearly non-existent, false-positive percentage.

SECTION OVERVIEW

Since GWAVA 4’s spam engine will auto-adjust, it is very important that GWAVA 4 is configured properly. In addition to the proper configuration, GWAVA 4’s spam engine must be trained with examples of good and bad messages.

This guide will show how to properly configure and train GWAVA 4’s spam engine to stop spam with a high catch rate and a low, nearly non-existent, false-positive percentage.

Preparation

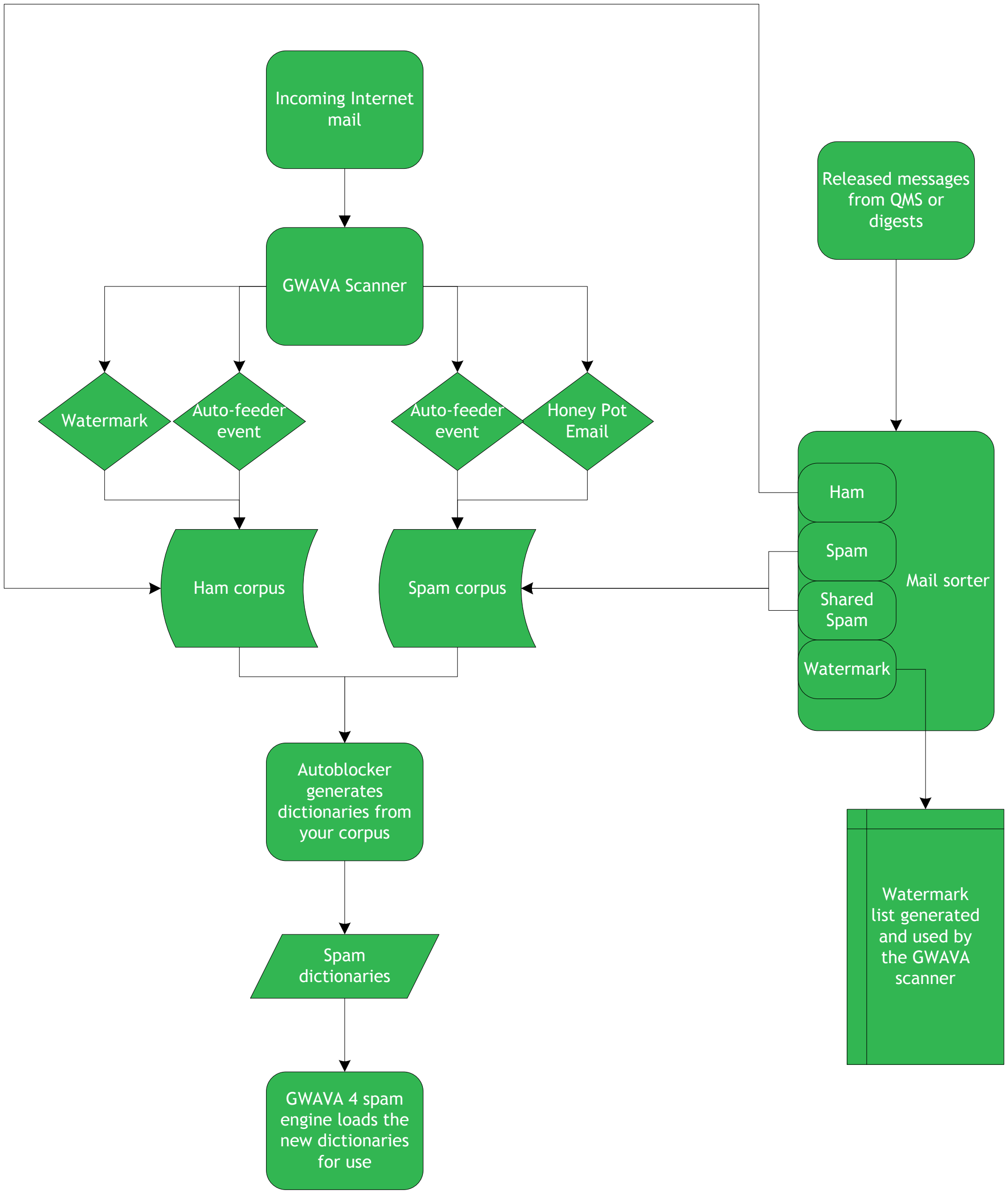
This section will walk through the steps required to create a mail sorter and preparing ham/spam guidelines for your organization. These guidelines are critical to the training process and must be followed closely to get the best results.

Pre-Training Configuration

Before training, GWAVA 4 will run in the default configuration mode. Until the training process has been completed GWAVA will use this configuration.

Training

The following diagram illustrates the training process.


Incoming Internet mail
GWAVA Scanner
Watermark
Auto-feeder event
Auto-feeder event
Honey Pot Email
Ham corpus
Spam corpus
Autoblocker generates dictionaries from your corpus
Spam dictionaries
GWAVA 4 spam engine loads the new dictionaries for use
Released messages from QMS or digests
Ham
Spam
Shared Spam
Watermark
Mail sorter
Watermark list generated and used by the GWAVA scanner

This guide outlines how to configure GWAVA to automatically train itself. In the diagram above the GWAVA Scanner is what will automatically add messages to the ham and spam corpus.

The other main training source will be the mail sorter (right). The mail sorter will be a dedicated mailbox that is used to manually sort messages to add to the ham and spam corpus. The training process will outline how to populate this mailbox and use it to train GWAVA.

After GWAVA is configured properly little user interaction will be necessary.

Your goal of the training process is to add mail to the ham and spam corpus. GWAVA will do the rest of the work.

Post-Training Configuration

GWAVA will automatically switch into a different configuration once you have supplied enough messages to the spam and ham corpora.

Appendices

The appendices show how to configure all of GWAVA's anti-spam options. Only use these options if you are not getting the desired results. You must have a high familiarity with GWAVA to use these options. Otherwise, it is highly recommended that you use the recommended configuration found in the post-training configuration section.

PREPARATION

Before proceeding to the Training section, (1) create ham/spam guidelines and (2) create a ham/spam mailbox. These preparation steps are not optional.

Ham/Spam Guidelines

Create a clear set of guidelines defining which email messages are ham and which are spam.

Spam is an unsolicited email message that is unwanted.

Spam is a message that is unsolicited and unwanted. Decide what is considered unwanted for your particular email system. Some examples of spam are stock messages, prescription medication advertisements, virus infected messages, company advertisements, and other unwanted messages.

Ham is a solicited email message that fulfills business purposes.

Ham is a solicited message that fulfills business purposes. Decide what is considered to be a message that fulfills business purposes. Some examples of ham are invoice messages, support messages, messages between employees, messages between companies, and other messages relevant to your business.

Once the guidelines are set for your company, follow the guidelines throughout the training process. Be consistent, and good results will follow. Please write the guidelines on the lines provided.

Ham:

(e.g. Personal messages, newsletters, student mail, etc.)

____

____

____

____

____

____

Notice that newsletters are listed in both examples of ham and spam; make the decision on which types of newsletters are acceptable for the company. For example Dell deals could be allowed but not newsletters from gamespot.com.

Spam:

(e.g. unsolicited mail, offensive mail, newsletters, etc.)

____

____

____

____

____

____

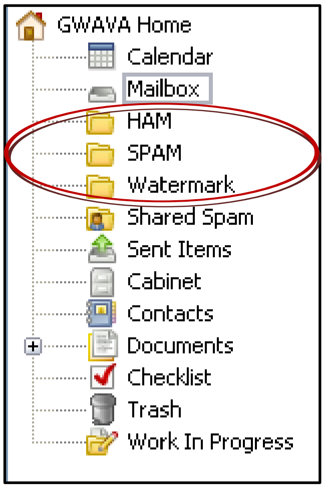
If the guidelines are not completely defined that’s OK. During the training process these guidelines become clearer.

Create Ham/Spam Mailbox

Create a dedicated mailbox for sorting ham and spam. GWAVA 4 uses the sorted messages as training sources. A large majority of the training will be done through this mailbox.

Create a mailbox for the user 'GWAVA' that will act as a sorter and assign it a password. Now create three folders at the root of the mailbox: HAM, SPAM, and Watermark.

After the mailbox is created and the folders are added, the mailbox's folder list should look similar to the following:



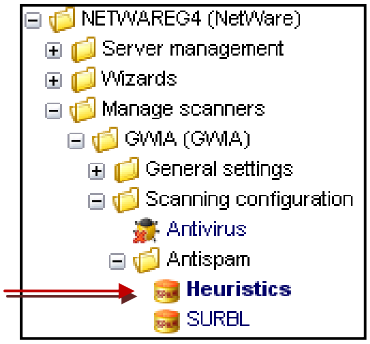
TIP: It is a good idea to use this mailbox's login and password for GWAVA's SMTP relay.

Make sure that the folders are located in the root of the mailbox.

Assign yourself proxy access to this mailbox for easy access.

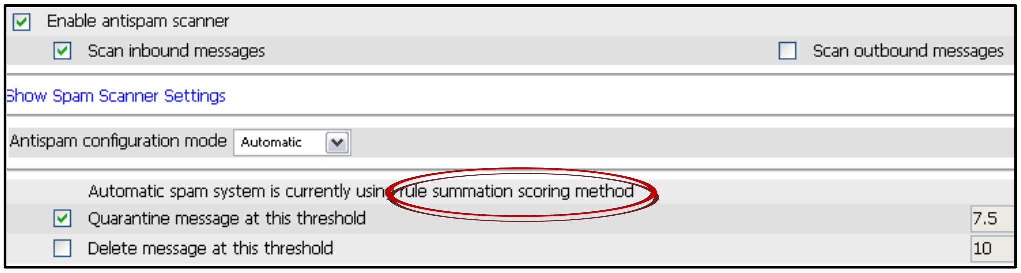
PRE-TRAINING CONFIGURATION

Before the GWAVA 4 probability engine is trained, it cannot be used. GWAVA will rely on normal heuristic scanning until the system is trained. After installing a scanner, open the scanner, and go to the 'Heuristics' section under the 'Antispam' scanning configuration.



By default after an install, GWAVA will be using basic heuristic scanning with a threshold of 7.5. Heuristic scanning is exactly how GWAVA 3.6 worked. This method will block a fair amount of spam with a low false positive percentage.

This configuration will not yield the best results. This configuration blocks some spam with a low false-positive rate. Use this configuration until you have completed the training process.



☑ Enable antispam scanner

☑ Scan inbound messages ☐ Scan outbound messages

Show Spam Scanner Settings

Antispam configuration mode: Automatic

Automatic spam system is currently using rule summation scoring method

☑ Quarantine message at this threshold 7.5

☐ Delete message at this threshold 10

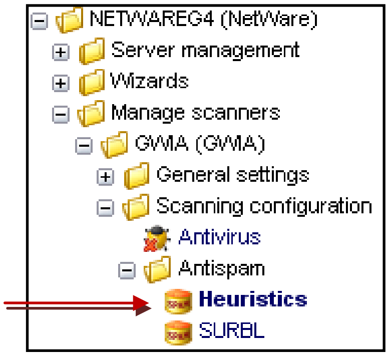
GWAVA 4 will continue to run in this mode until you have added at least 1000 ham messages and 1000 spam messages to the email corpus. This page will also tell you what mode it is using. The probability engine isn't trained yet, so the 'rule summation scoring method' is used.

TRAINING

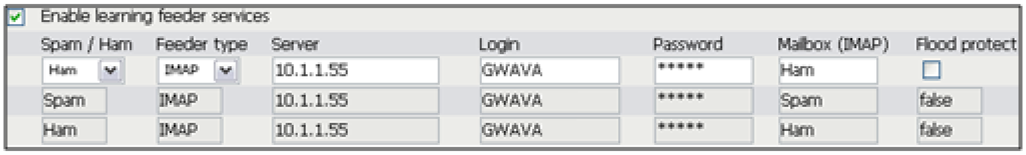
GWAVA4 trains the new spam engine based on examples of ham and spam. Ham and spam example gathering will be demonstrated in the following sections. After the samples are gathered, they will be classified through the mail sorter and then submitted for training.

Ham/Spam Feeder Configuration

Before sorting ham and spam, GWAVA must be configured to pull mail from the mailbox set up as the mail sorter. Remember this mailbox will simply be a convenient place to sort messages.



The feeder for GWAVA will automatically pull email from the ham and spam folders of the mail sorter. To configure this, click 'Enable Learning Feeder Services', then select 'Spam' or 'Ham' under 'Spam/Ham', enter the IP address of the post office under 'Server', under 'Login' enter the username of the mailbox, enter in the password, and under 'Mailbox(IMAP)' enter the folder name you created in your mailbox for either spam or ham. Then click the green plus sign and save the changes.



☑ Enable learning feeder services

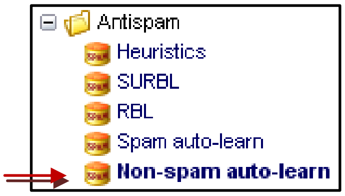
Spam / Ham	Feeder type	Server	Login	Password	Mailbox (IMAP)	Flood protect
Ham	IMAP	10.1.1.55	GWAVA	*****	Ham	□
Spam	IMAP	10.1.1.55	GWAVA	*****	Spam	false
Ham	IMAP	10.1.1.55	GWAVA	*****	Ham	false

Notice that the 'Flood protect' option has been left blank. To control the amount of ham or spam that is learned by the probability engine, GWAVA 4 uses flood protection. Leaving this option unchecked allows all messages that are manually sorted to be accepted for training.

See the flood protection section for details.

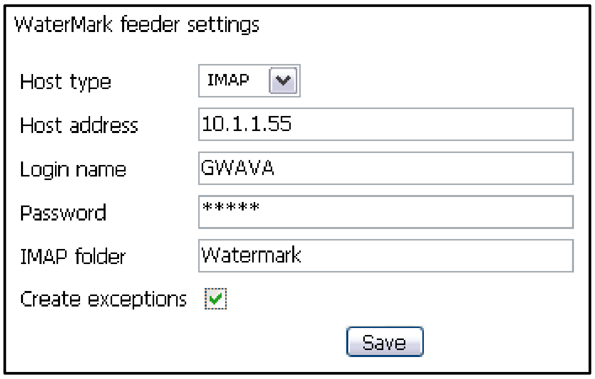
Watermark Feeder Configuration

GWAVA needs to be configured so it will automatically pull messages from the watermark folder of the mail sorter. Go to the ‘Non-spam auto-learn’ section under the ‘Antispam’ configuration.



Click ‘Watermark Quick-Feeder’ and then click ‘add a new training source’. Fill out the information that connects to the mail sorter and the watermark folder. Check ‘Create Exceptions’ and that will make sure that watermarked messages never get blocked again by the spam engine.

See the watermarking section for details.



WaterMark feeder settings

Host type	IMAP
Host address	10.1.1.55
Login name	GWAVA
Password	*****
IMAP folder	Watermark
Create exceptions	☑

GWAVA will now check every 10 minutes for messages to retrieve from the Watermark folder.

Mail Sorter

Population of the mail sorter will be discussed in the ham feeding and spam feeding sections.

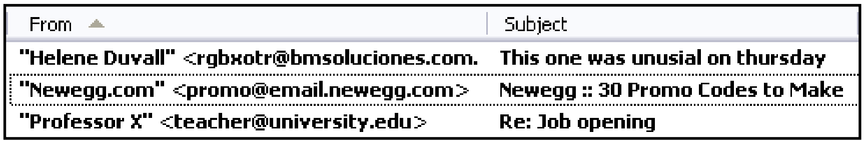
The mail sorter is the central hub of the training process. The majority of your time will be spent going through this mailbox.

Example messages can be sorted into the folders created. The messages will then be pulled from those folders by GWAVA4 and used to train the spam engine.

Remember to follow the ham/spam guidelines.

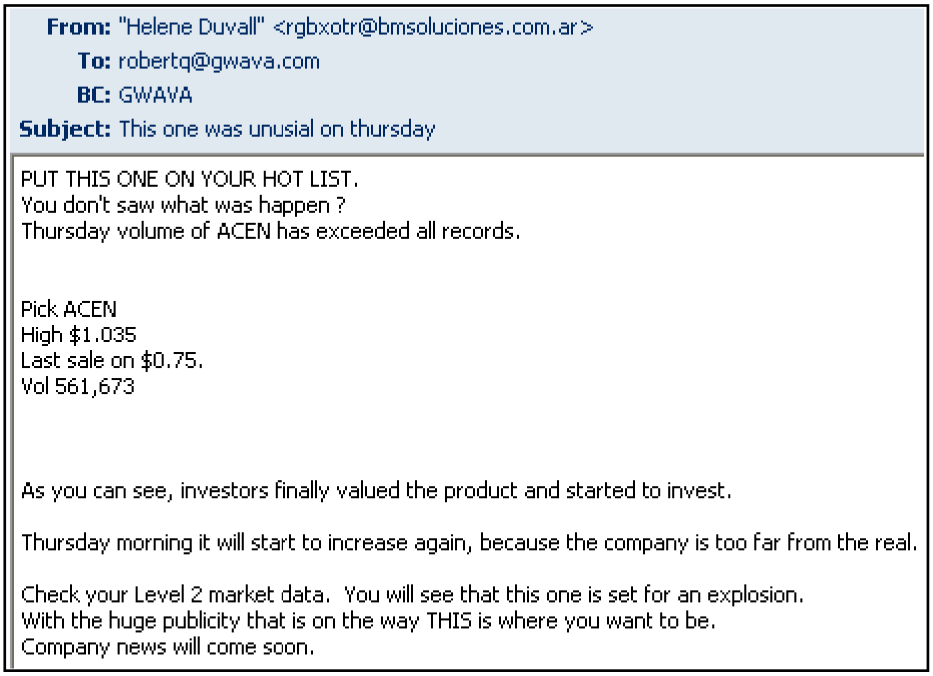
If a message is spam, drop it into the spam directory. If a message is ham, drop it into the ham directory. If a message should be watermarked, drop it into the watermark folder.

As an example, some messages are in an example mailbox to illustrate how to use the mail sorter. The mailbox contains three messages; each message must be reviewed to see which folder they should be in.



From	Subject
"Helene Duvall" <rgbxotr@bmsoluciones.com.	This one was unusial on thursday
"Newegg.com" <promo@email.newegg.com>	Newegg :: 30 Promo Codes to Make
"Professor X" <teacher@university.edu>	Re: Job opening

Open the email from Helene.



From: "Helene Duvall" <rgbxotr@bmsoluciones.com.ar>
To: robertq@gwava.com
BC: GWAVA
Subject: This one was unusial on thursday

PUT THIS ONE ON YOUR HOT LIST.
You don't saw what was happen ?
Thursday volume of ACEN has exceeded all records.

Pick ACEN
High $1.035
Last sale on $0.75.
Vol 561,673

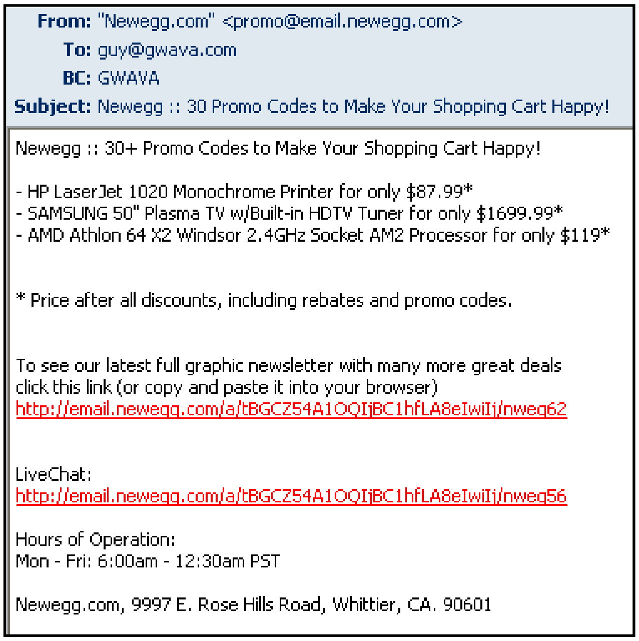
As you can see, investors finally valued the product and started to invest.

Thursday morning it will start to increase again, because the company is too far from the real.

Check your Level 2 market data. You will see that this one is set for an explosion.
With the huge publicity that is on the way THIS is where you want to be.
Company news will come soon.

Right away it is obvious that this message is spam; promptly drop it into the spam folder for GWAVA to retrieve.

Open the second email from NewEgg.com.



From: "Newegg.com" <promo@email.newegg.com>
To: guy@gwava.com
BC: GWAVA
Subject: Newegg :: 30 Promo Codes to Make Your Shopping Cart Happy!

Newegg :: 30+ Promo Codes to Make Your Shopping Cart Happy!

- HP LaserJet 1020 Monochrome Printer for only $87.99*
- SAMSUNG 50" Plasma TV w/Built-in HDTV Tuner for only $1699.99*
- AMD Athlon 64 X2 Windsor 2.4GHz Socket AM2 Processor for only $119*

* Price after all discounts, including rebates and promo codes.

To see our latest full graphic newsletter with many more great deals
click this link (or copy and paste it into your browser)
http://email.newegg.com/a/tBGCZ54A1OQIjBC1hfLA8eIwiIj/nweg62

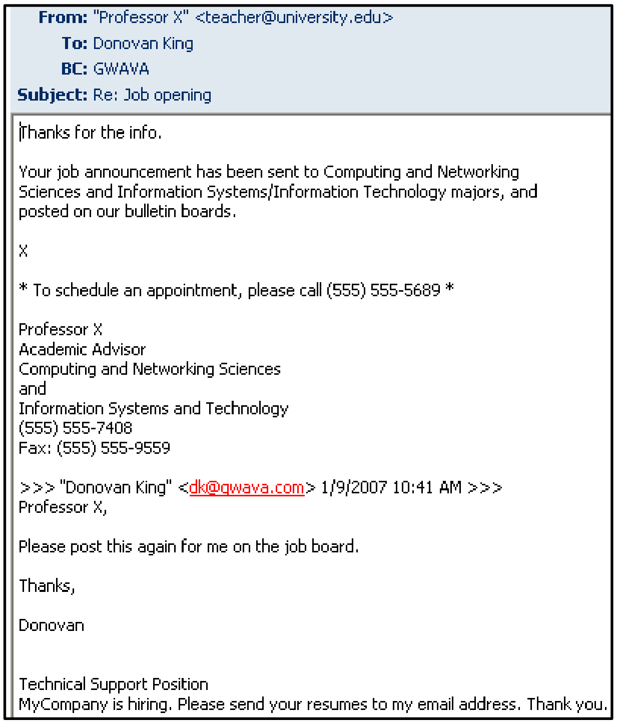
LiveChat:
http://email.newegg.com/a/tBGCZ54A1OQIjBC1hfLA8eIwiIj/nweg56

Hours of Operation:
Mon - Fri: 6:00am - 12:30am PST

Newegg.com, 9997 E. Rose Hills Road, Whittier, CA. 90601

This message is a newsletter from NewEgg.com. NewEgg is allowed in this mail environment, so drop the message into the ham folder.

Open the third message from Professor X.



From: "Professor X" <teacher@university.edu>
To: Donovan King
BC: GWAVA
Subject: Re: Job opening

Thanks for the info.

Your job announcement has been sent to Computing and Networking Sciences and Information Systems/Information Technology majors, and posted on our bulletin boards.

X

* To schedule an appointment, please call (555) 555-5689 *

Professor X
Academic Advisor
Computing and Networking Sciences
and
Information Systems and Technology
(555) 555-7408
Fax: (555) 555-9559

>>> "Donovan King" <dk@gwava.com> 1/9/2007 10:41 AM >>>
Professor X,

Please post this again for me on the job board.

Thanks,

Donovan

Technical Support Position
MyCompany is hiring. Please send your resumes to my email address. Thank you.

This message is also ham, but it happens to be a message that qualifies for a watermark. Drop the message into the watermark folder.

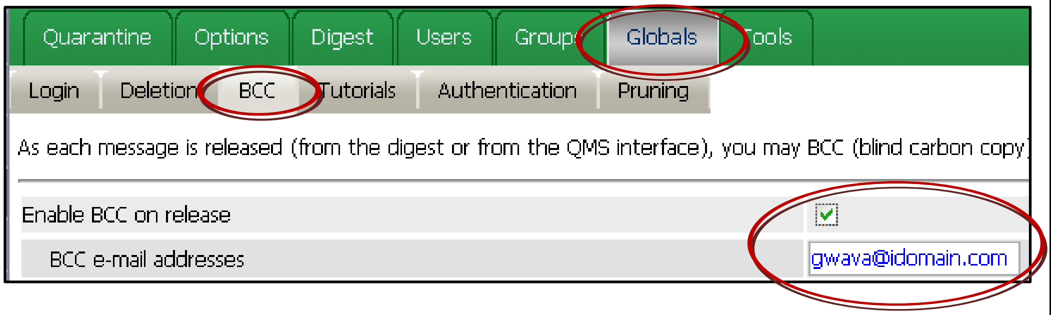
Ham Feeding

This section will outline some of the best methods to use for getting examples of ham into the mail sorter and how to configure GWAVA to train itself off of those messages.

Resubmissions BCC Configuration

The easiest way to get ham samples is for released messages to get dropped into the mail sorter. To do this open QMS (http://<ip_address>:49285), log in as administrator.

Navigate to the 'Globals' tab and then click the 'BCC' subtab. Put a checkmark in 'Enable BCC on release' and enter in the email address of the mail sorter.



As each message is released (from the digest or from the QMS interface), you may BCC (blind carbon copy

Enable BCC on release ☑

BCC e-mail addresses gwava@idomain.com

You should have plenty of mail in this mailbox to sort during the first week or so. As time goes on the amount of ham in there should be minimal. Due to your training, there will be very little spam that makes it through the system without blocking ham.

Check this mailbox frequently for new training sources, and drop the messages in the appropriate folder.

The guidelines for watermarks are explained in the watermarking section.

For this method to be most effective digests should be enabled.

Not all released messages are HAM. Make sure you review them in your mail sorter, before you add them to the HAM corpus.

All released messages from QMS will be BCC'd to the mail sorter. After a few minutes, messages will show up in the mail sorter. Sort the messages by dropping ham into the ham folder and spam into the spam folder.

Watermarking

A watermark is a special type of ham message that is very important in the training process. A message that is watermarked will be added into the ham corpus, and any future messages from that person will be added to the ham corpus automatically. If the exception option is enabled for the watermarks, future message will not be scanned for spam content. Here are some basic guidelines for a ham message that should be watermarked.

Don't worry. When a message has been watermarked it won't add any messages to the ham corpus in the future that have spoofed senders.

- The message should be one on one communication (no mass-mailings)
- Be certain the sender will never send spam

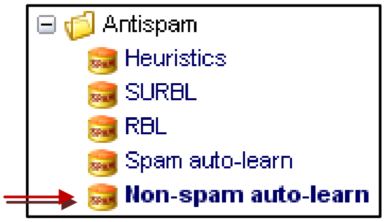
Mail Sorter Watermarking

As messages show up in the mail sorter, be on the lookout for messages that can be watermarked. With enough watermarks GWAVA can learn ham automatically and released messages will show up less and less in the mail sorter.

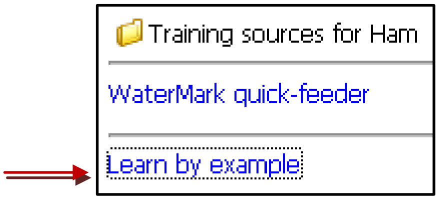
If the message came from Yahoo, Gmail, AOL, MSN, or Hotmail, you may have to add messages from that sender as a watermark several times.

Manual Watermarking

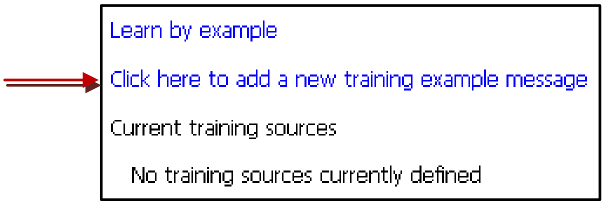
To manually create watermarks, browse to the 'Non-spam auto-learn' section in GWAVAMAN.



Now click on the 'Learn by Example' link.



To get started, click 'add a new training example message'.



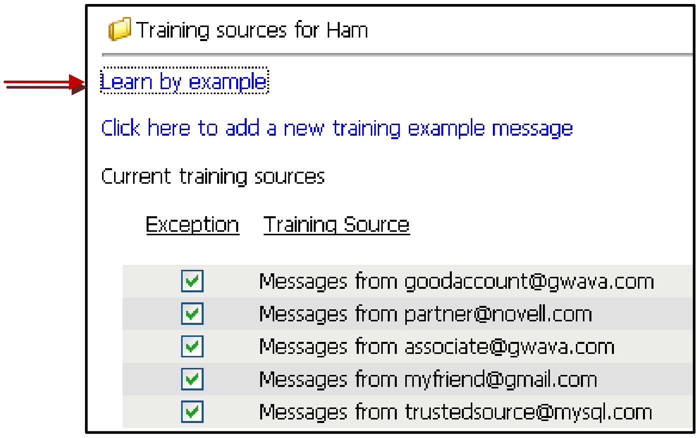
Current training sources

No training sources currently defined

From this screen you can upload a MIME file that has been previously saved or paste the contents of a MIME file.

Once a file has been uploaded, make sure that these messages are never blocked by the anti-spam system by checking 'Exclude this sender from spam scanning'. Create a label for the watermark. Then click the ‘Submit Watermark’ button. To submit another example, click the link to go through the process again.

Added examples can be reviewed by going back to the initial watermark page and reviewing the list of watermarks.



Training sources for Ham

Learn by example

Click here to add a new training example message

Current training sources

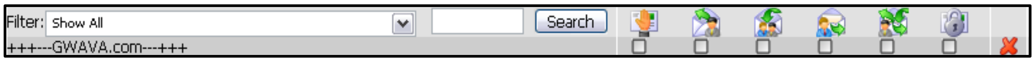
Exception	Training Source
☑	Messages from goodaccount@gwava.com
☑	Messages from partner@novell.com
☑	Messages from associate@gwava.com
☑	Messages from myfriend@gmail.com
☑	Messages from trustedsource@mysql.com

Delete accidental entries by clicking the ✗ .

Ham Auto-Learning

In the ‘Non-spam auto-learn’ section, certain events can be automatically added to the ham corpus.

One possibility is to create a text filter that contains specific text from your company signature. When creating this filter be sure **NOT** to associate any services (block, quarantine, etc.) with the message, because we want the message to be delivered.

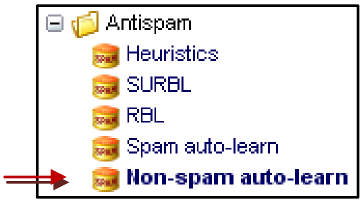


Filter: Show All

+++---GWAVA.com---+++

The only reason to create the filter is so that it would show up in the event list for ham auto-learning.

To add the auto-learning event, open the ‘Non-spam auto-learn’ section of GWAVAMAN.



For messages in your mailbox, view the source of messages in your mailbox and copy and paste the source MIME information to use for a watermark.

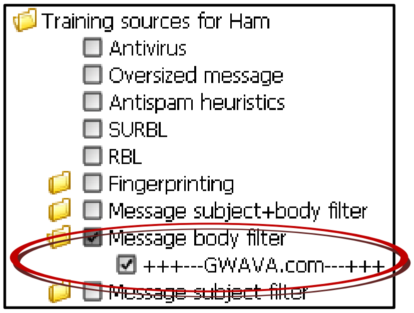
Watermarks made with the quick-feeder will also show up in this list.

If any messages in this list have a corrupted sender, delete them.

Most events should not be added automatically to the ham corpus.

If you choose to do this, make sure that the word you use is not part of your internet domain and only appears in ham messages. Do not use words that appear in normal sentences.

Open the ‘Training sources for Ham’ folder. I created a body filter so I click on the folder for ‘Message body filter’ and check ‘+++---GWAVA.com---+++’. When the filter fires on a message, that message will be added to the ham corpus.



Training sources for Ham
- ☐ Antivirus
- ☐ Oversized message
- ☐ Antispam heuristics
- ☐ SURBL
- ☐ RBL
- ☐ Fingerprinting
- ☐ Message subject+body filter
- ☑ Message body filter
  - ☑ +++---GWAVA.com---+++
- ☐ Message subject filter

When someone responds to an email containing your company signature, that email will be added to the ham corpus.

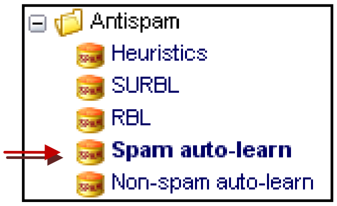
Spam Feeding

This section will outline some of the best methods to use for getting examples of spam into the mail sorter and how to configure GWAVA to train itself off of those messages.

You could create some text filters on disallowed words that could be automatically fed.

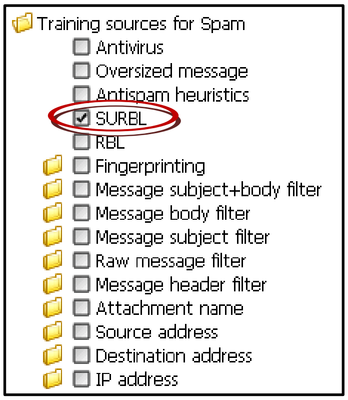
Spam Auto-Learning

To automatically add events to the spam corpus go to the ‘Spam auto-learn’ section.



Be sure the event fires on spam 100% of the time.

After opening up the ‘Training sources for Spam’ folder, a list of all the events that could fire for a message is displayed. From here choose events that will fire on spam. When a message comes into the system and the selected event fires, it will automatically be added to the spam corpus.



Training sources for Spam
- ☐ Antivirus
- ☐ Oversized message
- ☐ Antispam heuristics
- ☑ SURBL
- ☐ RBL
- ☐ Fingerprinting
- ☐ Message subject+body filter
- ☐ Message body filter
- ☐ Message subject filter
- ☐ Raw message filter
- ☐ Message header filter
- ☐ Attachment name
- ☐ Source address
- ☐ Destination address
- ☐ IP address

Check SURBL. SURBL will almost always be spam.

Do **NOT** check RBL, because it is too susceptible to false positives.

All SURBL messages will get added to the spam corpus automatically in this example.

Honey Pot Setup/Feeding

Spammers begin their spam attack by obtaining a list of mailboxes that exist in your email system. They obtain a list of mailboxes by using a dictionary attack or even by guessing common mailbox names. Spammers will then use this list to send spam to your email system via a script of some sort. Honey pot email accounts are email accounts that are set up to collect these unsolicited messages from spammers.

Honey pot email accounts should never receive legitimate email.

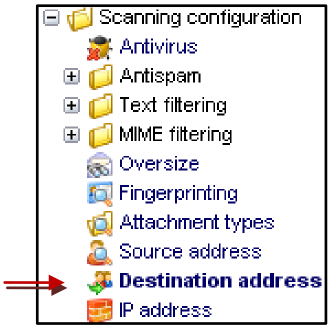
Spam caught in honey pots is an excellent training source for GWAVA4. GWAVA 4 can automatically add messages sent to honey pots to the spam corpus. This section will outline how to configure GWAVA to utilize honey pots.

First, create five or six honey pots. Create accounts like marketing@domain.com, info@domain.com, sales@domain.com, accounting@domain.com, hr@domain.com, postmaster@domain.com, webmaster@domain.com, and shipping@domain.com; just to name a few possibilities. These accounts should be tempting targets. Spammers will discover these accounts exist and will start sending spam to them.

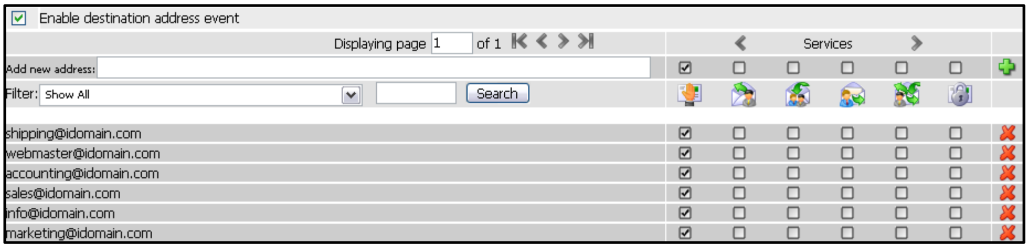
Before creating these accounts, make sure they do not exist and will never exist.

Second, create destination address blocks for each account. Go to the destination address block section inside the web configuration utility.

These addresses could be set up as GroupWise resources instead of normal GroupWise users.



Add each honey pot account to the list, and check the block service for each account.



☑ Enable destination address event

Displaying page 1 of 1

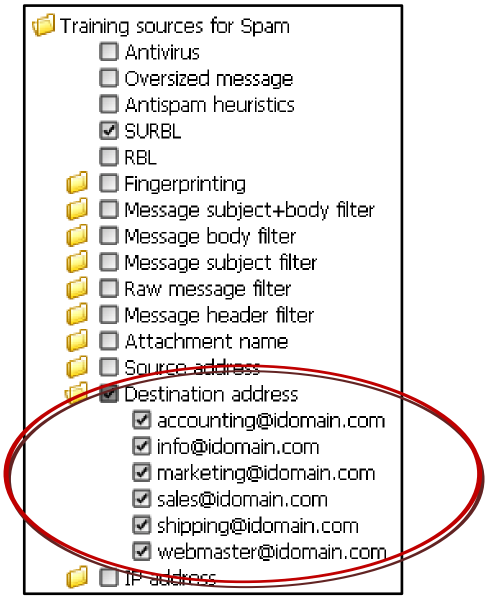
Services

Add new address:

Filter: Show All Search

Add new address:	☑	□	□	□	□	□
shipping@idomain.com	☑	□	□	□	□	□
webmaster@idomain.com	☑	□	□	□	□	□
accounting@idomain.com	☑	□	□	□	□	□
sales@idomain.com	☑	□	□	□	□	□
info@idomain.com	☑	□	□	□	□	□
marketing@idomain.com	☑	□	□	□	□	□

Lastly, open the 'Spam auto-learn' section from the 'Heuristics' folder and check off the destination address events to automatically be added to the SPAM corpus.



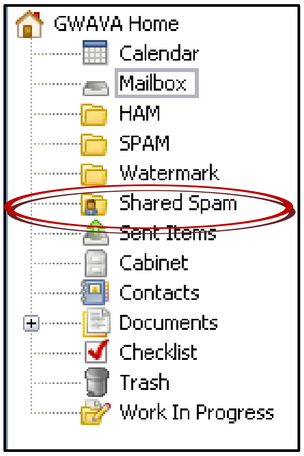
Training sources for Spam
- □ Antivirus
- □ Oversized message
- □ Antispam heuristics
- ☑ SURBL
- □ RBL
- □ Fingerprinting
- □ Message subject+body filter
- □ Message body filter
- □ Message subject filter
- □ Raw message filter
- □ Message header filter
- □ Attachment name
- □ Source address
- ☑ Destination address
  - ☑ accounting@idomain.com
  - ☑ info@idomain.com
  - ☑ marketing@idomain.com
  - ☑ sales@idomain.com
  - ☑ shipping@idomain.com
  - ☑ webmaster@idomain.com
- □ IP address

Now messages that are sent to the honey pots will automatically be added to the spam corpus.

Shared Spam Folder

Spam messages that get past GWAVA to user mailboxes are also a good training source.

Create a shared spam folder and send invitations to trusted co-workers.



GWAVA Home
- Calendar
- Mailbox
- HAM
- SPAM
- Watermark
- Shared Spam
- Sent Items
- Cabinet
- Contacts
- Documents
- Checklist
- Trash
- Work In Progress

Instruct users with access to the shared folder to drop any spam messages that got through to their mailbox into this folder.

After users drag messages to the shared folder, verify all of the messages are spam in your mail sorter, and drop them into the spam folder for training.

Only check the destination address blocks created for the honey pots.

Honey pots are highly recommended so that any future types of spam can be learned and blocked in a timely manner.

For details on how to set up a shared folder refer to your mail system's documentation.

Use this method with trusted users. Don't give access to the shared folder to the entire company.

Make sure your mail sorter has access to the shared spam folder.

Flood Protection

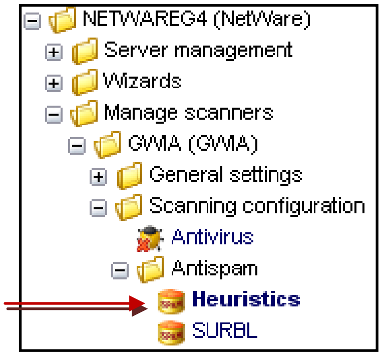
Flood protection limits the amount of mail that can be automatically learned by probability engine. Flood protection creates a quota of mail that needs to be added to the ham/spam corpus. About 2 messages of ham and 2 messages of spam are expected each minute. If you have 100 spam messages added in one minute from SURBL hits, it would only train off of 2 of them per minute and discard the rest. The same would be true of email that is added from the ham auto-feeders. However, if for 5 minutes no mail was transferred, the quota would then be 12 messages for the next minute, and it could take up to that many messages.

The spam/ham auto-learning functions activate flood protection automatically. For mail feeders, you control whether flood protection is enabled. The administrator will not want to have flood protection enabled for your mail feeders. All mail sorted using the mail sorter should be added to the corpus.

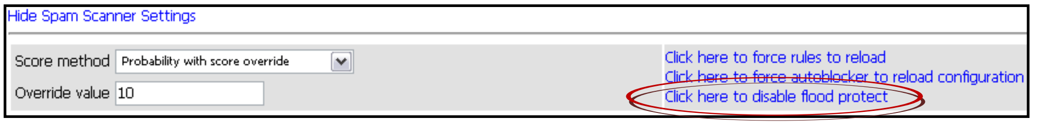
Manually Train Engine

Some administrators may already have collected a spam/ham corpus and want to import the corpus manually into the GWAVA system. Please go through your spam/ham corpus and fix any obvious classification errors before doing this.

Flood protection must be disabled for this training method to work properly. To disable flood protection go to the 'Antispam' heuristics page.



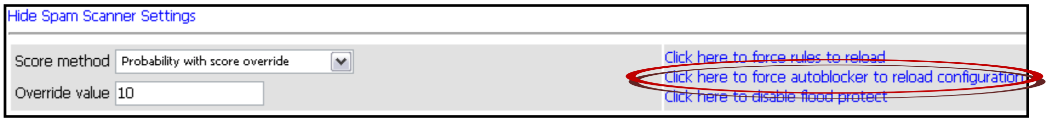
Click 'Show Spam Scanner Settings'. Then click 'disable flood protect'.


Hide Spam Scanner Settings
Click here to force rules to reload
Click here to force autoblocker to reload configuration
Click here to disable flood protect

Score method Probability with score override

Override value 10

The configuration changes need to be loaded by Autoblocker. Click 'force autoblocker to reload configuration' for the changes to take effect.


Hide Spam Scanner Settings
Click here to force rules to reload
Click here to force autoblocker to reload configuration
Click here to disable flood protect

Score method Probability with score override

Override value 10

To import the files, drop the MIME files into the <*GWAVA Install Folder>/services/autoblocker/transfer/<interface_ID>/[Ham or Spam]* folder. GWAVA will automatically use these files to train itself.

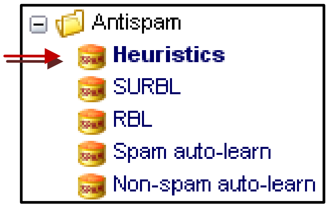
After dropping in the messages, turn flood protection back on by clicking ‘enable flood protect’. Then reload the Autoblocker configuration by clicking ‘force autoblocker to reload configuration’.

Training is a continual process. Continue using the methods discussed to continually train the GWAVA 4 spam engine.

POST-TRAINING CONFIGURATION

After you have given the system 1,000 ham examples and 1,000 spam examples the configuration will switch to the probability with score override.

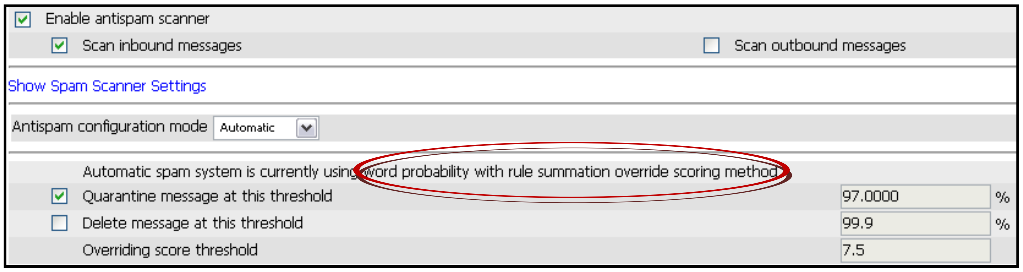
Go to the heuristics section of the scanner configuration.



Probability with score override will begin to use the training you have given the system to block more spam.

Now GWAVA uses both the old heuristics engine and the new probability engine. If the new probability engine is 97% sure the message is SPAM it will block the message and quarantine it, and if the old engine gives the message a score of 7.5 or above, the message will also be blocked or quarantined.

Continue training the system to get to the final scanning mode.



☑ Enable antispam scanner

☑ Scan inbound messages ☐ Scan outbound messages

Show Spam Scanner Settings

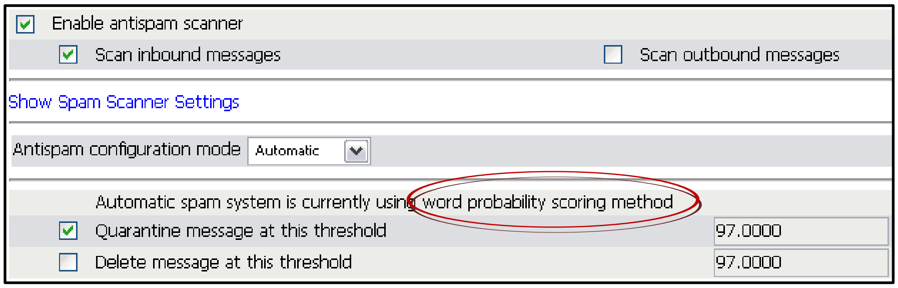
Antispam configuration mode: Automatic

Automatic spam system is currently using word probability with rule summation override scoring method

☑ Quarantine message at this threshold	97.0000	%
☐ Delete message at this threshold	99.9	%
Overriding score threshold	7.5	

This is just an interim step in the automatic configuration; once you have given the system 5,000 ham samples and 10,000 spam samples the automatic configuration will automatically switch to probability only scanning.

The final step in the automatic configuration is the ‘word probability’ method.



☑ Enable antispam scanner

☑ Scan inbound messages ☐ Scan outbound messages

Show Spam Scanner Settings

Antispam configuration mode: Automatic

Automatic spam system is currently using word probability scoring method

☑ Quarantine message at this threshold	97.0000
☐ Delete message at this threshold	97.0000

GWAVA should now be getting great results, and as time goes on GWAVA will get better and better.

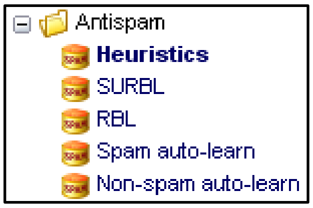
More information about the heuristic engine and configuration modes can be found in the appendices.

APPENDIX A: SCANNING MODES

This section outlines some of the major configuration modes that can be used in GWAVA 4.

Automatic Configuration Mode

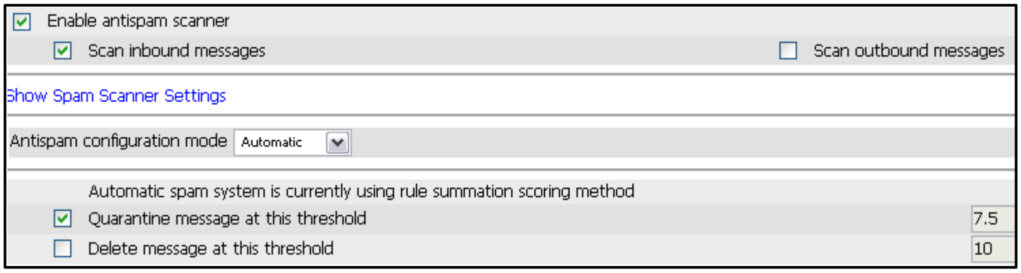
By default, GWAVA starts spam scanning in the 'Automatic Configuration mode' which uses the heuristic spam engine.



Antispam
- **Heuristics**
- SURBL
- RBL
- Spam auto-learn
- Non-spam auto-learn

The scores cannot be set while in this mode. GWAVA uses a hardcoded score of 7.5, which should block a reasonable amount of spam without having any false positives. The results at this stage will not be the best, but only serve as an interim until the administrator can switch over to the newer spam engine.

See heuristic scanning for more information.



☑ Enable antispam scanner

☑ Scan inbound messages □ Scan outbound messages

Show Spam Scanner Settings

Antispam configuration mode: Automatic

Automatic spam system is currently using rule summation scoring method

☑ Quarantine message at this threshold 7.5

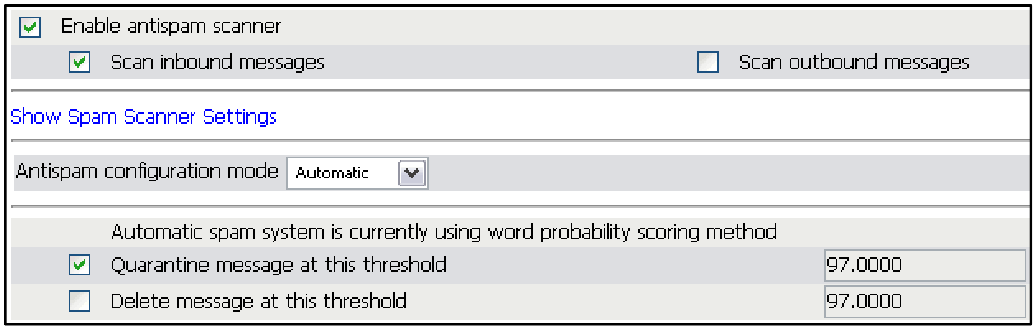
□ Delete message at this threshold 10

GWAVA will use the heuristic scanning engine until you have fed GWAVA 1000 examples of ham and 1000 examples of spam. Once you have the appropriate set of ham and spam, the anti-spam engine will automatically switch over to the new engine, which is based on probability.

See the training section for more information on gathering ham and spam.

How GWAVA calculates the probability is somewhat complicated, but essentially GWAVA looks at the word content of the message and gives you a spam percentage based on what it has learned from your examples of ham and spam. After GWAVA switches over you will see a screen that resembles the following.

See probability scanning for more information.



☑ Enable antispam scanner

☑ Scan inbound messages □ Scan outbound messages

Show Spam Scanner Settings

Antispam configuration mode: Automatic

Automatic spam system is currently using word probability scoring method

☑ Quarantine message at this threshold 97.0000

□ Delete message at this threshold 97.0000

Notice that the spam percentage is given instead of an arbitrary spam score. This means that the probability engine blocks and quarantines the message if it is 97% sure the message is spam.

The thresholds are not editable unless you switch to a different configuration mode.

Simple Configuration Mode

Simple configuration mode is easy to use and resembles how GWAVA 3.6 worked, except that you can choose which engine(s) to use.

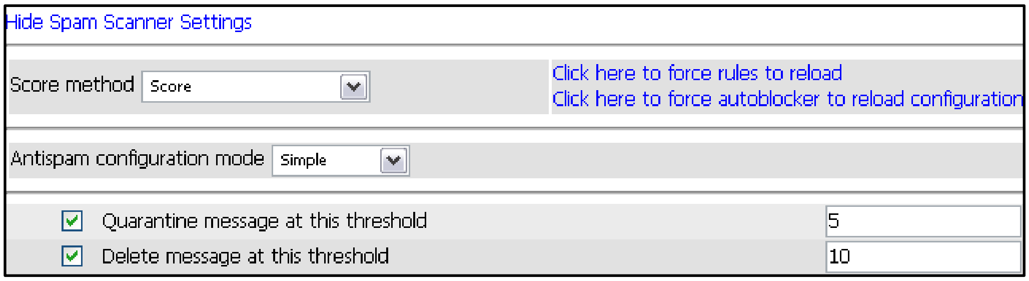
Heuristic Scanning

Heuristic scanning is similar to the spam scanning that was done in GWAVA 3.6.

With heuristic scanning, a series of pre-compiled rules (mostly regular expressions) are executed against the content of the messages and their headers. Each rule is assigned a score value, and the total score is the summation of the scores of the rules that fired.

A starting PCR file is provided with your GWAVA installation. Your old GWAVA 3.6 rule file may be used if you have made several rules or have done extensive optimization. If you would like to import your PCR file from GWAVA 3.6 you can copy your old PCR file from your NetWare server <groupwise_domain>\gwava\config\spamcfg\compiled.pcr to <GWAVAA4_Install_Folder>/services/asengine/configs/<interface_ID>/ on your GWAVA4 server.

To use heuristic scanning you must browse to the Heuristics page inside GWAVAMAN. Set the 'Antispam configuration mode' to 'Simple'. Then click on 'Show spam scanner settings' and make sure that 'Score' is selected for the Score Method.



Hide Spam Scanner Settings

Score method: Score

Click here to force rules to reload
Click here to force autoblocker to reload configuration

Antispam configuration mode: Simple

☑ Quarantine message at this threshold: 5
☑ Delete message at this threshold: 10

Quarantine at this threshold - all messages exceeding this score (but below the Delete message at this threshold score) will be blocked and quarantined.

Delete message at this threshold - all messages exceeding this score will be blocked and will **NOT** be put in the quarantine.

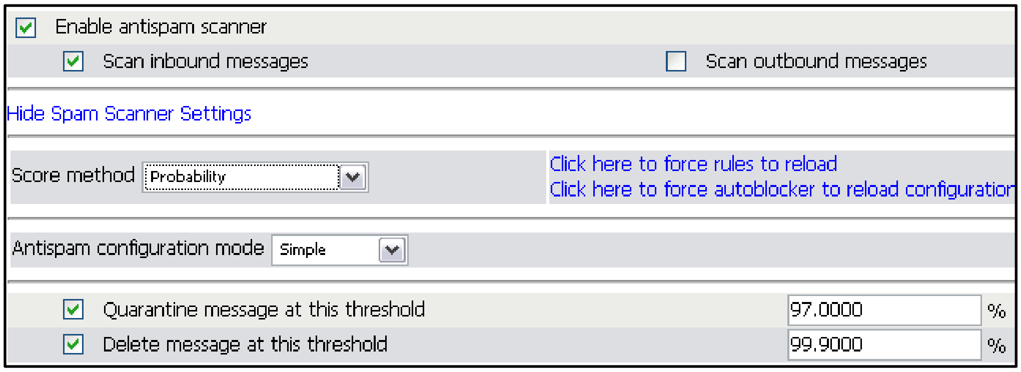
Probability Scanning

This is the new spam engine that will be used by GWAVA once you have provided 1000 samples of ham and 1000 samples of spam to the anti-spam engine. **DO NOT** use probability mode until you have a corpus of 1000 ham and 1000 spam, because nothing will be blocked.

See the training section for more information on gathering ham and spam.

Probability scanning returns a spam probability (from 0% to 100%), rather than a score. The percentage returned reflects closely the ham and spam that you have used to train the engine.

To use probability scanning, browse to the Heuristics page inside GWAVAMAN. Set the 'Antispam configuration mode' to 'Simple'. Then click on 'Show spam scanner settings' and make sure that Probability is selected for the 'Score Method'.



☑ Enable antispam scanner

☑ Scan inbound messages ☐ Scan outbound messages

Hide Spam Scanner Settings

Score method: Probability

Click here to force rules to reload

Click here to force autoblocker to reload configuration

Antispam configuration mode: Simple

☑ Quarantine message at this threshold: 97.0000 %

☑ Delete message at this threshold: 99.9000 %

These thresholds can be lowered. The amount of spam caught will increase but false positives will also increase.

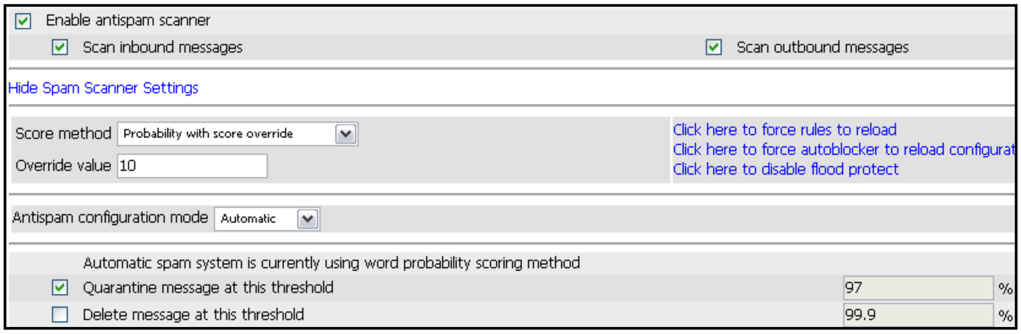
Quarantine at this threshold - 97% is your default threshold. The anti-spam engine blocks and quarantines messages when it is 97% sure the message is spam.

Delete message at this threshold - 99.9% is your default threshold. The anti-spam engine blocks messages when it is 99.9% sure that the message is spam. The message will not be quarantined.

Probability/Score with Override

'Probability with score override' and 'Score with probability override' are very similar to each other. They use both engines to determine if the message should be blocked. The method listed first will be your main engine and will be used for all of your normal thresholds. The second engine listed is your secondary engine.

The override value is the maximum allowed value of the secondary engine.



☑ Enable antispam scanner

☑ Scan inbound messages ☑ Scan outbound messages

Hide Spam Scanner Settings

Score method: Probability with score override

Override value: 10

Click here to force rules to reload

Click here to force autoblocker to reload configurat

Click here to disable flood protect

Antispam configuration mode: Automatic

Automatic spam system is currently using word probability scoring method

☑ Quarantine message at this threshold: 97 %

☐ Delete message at this threshold: 99.9 %

In this example the main engine is the probability engine and the secondary engine is the heuristics (score) engine.

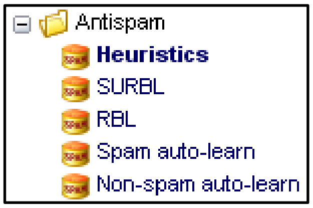
If the probability engine is 97% positive the message is SPAM the message will be blocked, and if the secondary heuristic engine returns a score greater than 10, the message will be blocked.

The reverse would be true for Score with probability override.

APPENDIX B: ADVANCED CONFIGURATION

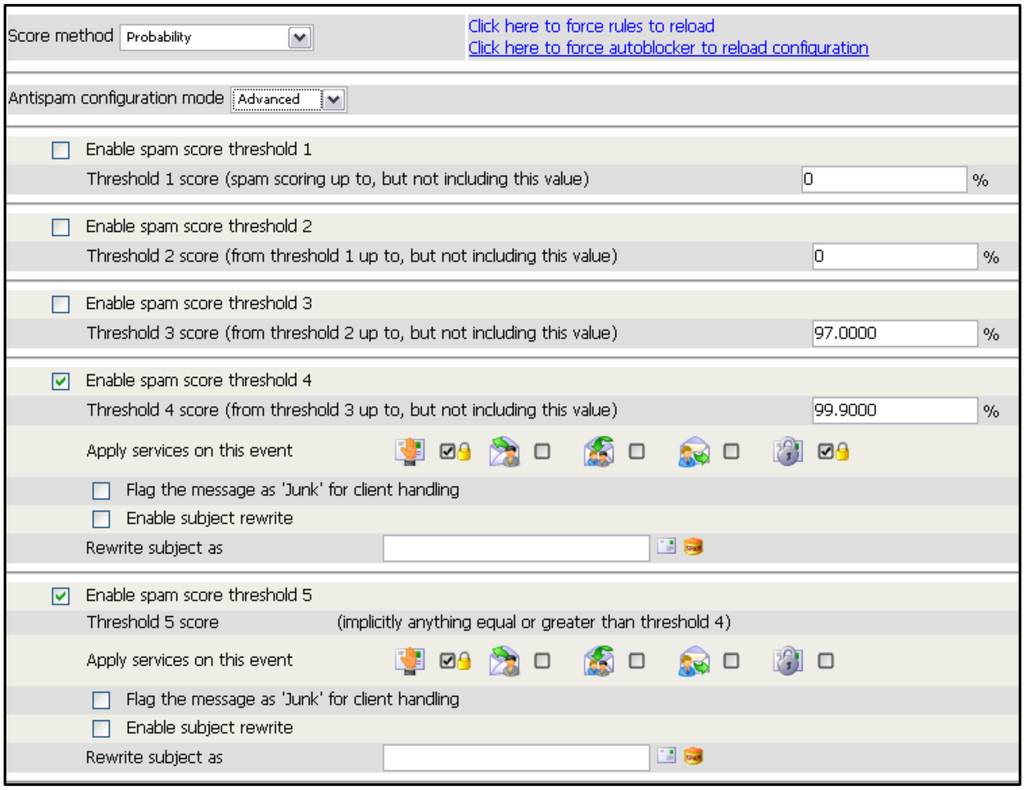
The advanced configuration mode allows you to have multiple thresholds and provides options to rewrite the subject or add headers for additional filtering with the mail client.

To use the advanced configuration mode you must browse to the Heuristics page inside GWAVAMAN.



- Antispam
  - **Heuristics**
  - SURBL
  - RBL
  - Spam auto-learn
  - Non-spam auto-learn

Set the ‘Antispam configuration mode’ to ‘Advanced’. GWAVA provides thresholds as either a scores or a probability percentage, depending the selected ‘Score method’.



Score method: Probability

Click here to force rules to reload
Click here to force autoblocker to reload configuration

Antispam configuration mode: Advanced

□ Enable spam score threshold 1
Threshold 1 score (spam scoring up to, but not including this value): 0 %

□ Enable spam score threshold 2
Threshold 2 score (from threshold 1 up to, but not including this value): 0 %

□ Enable spam score threshold 3
Threshold 3 score (from threshold 2 up to, but not including this value): 97.0000 %

☑ Enable spam score threshold 4
Threshold 4 score (from threshold 3 up to, but not including this value): 99.9000 %
Apply services on this event: ☑ □ □ □ ☑
□ Flag the message as 'Junk' for client handling
□ Enable subject rewrite
Rewrite subject as:

☑ Enable spam score threshold 5
Threshold 5 score (implicitly anything equal or greater than threshold 4)
Apply services on this event: ☑ □ □ □ □
□ Flag the message as 'Junk' for client handling
□ Enable subject rewrite
Rewrite subject as:

Configuring Thresholds

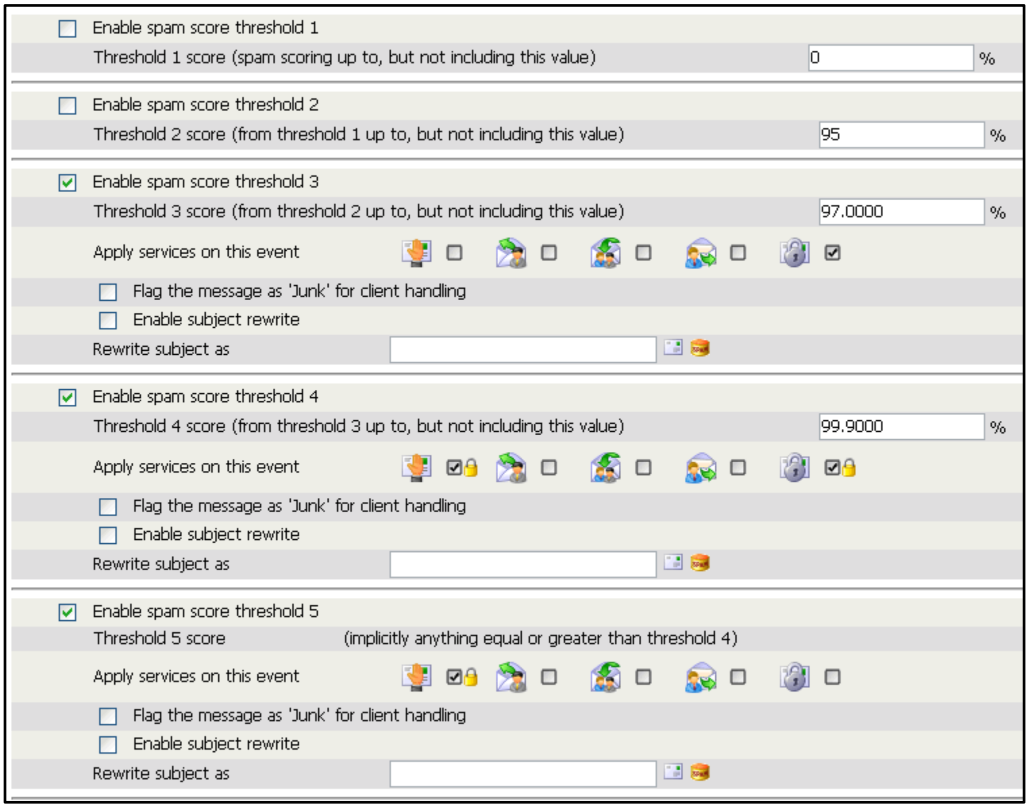
The advantage of having multiple thresholds is the administrator can set up different services to be activated at different score thresholds in a more granular fashion than permitted by the simple configuration mode.

GWAVA has the capability of configuring up to five threshold levels.

One possible configuration that uses multiple thresholds is:

- Messages below a spam probability of 95% will pass through GWAVA unaltered.
- Messages returning a spam probability of 95%-97% will be quarantined for review purposes. These messages are probably spam, but the administrator is being conservative regarding false positives. The original message will be delivered to the user.
- Messages returning a 97%-99.9% probability will be both blocked and quarantined. These messages are almost certainly spam.
- Messages exceeding a 99.9% probability will be blocked, but not quarantined.

Here is how this scenario could be configured:



□ Enable spam score threshold 1
Threshold 1 score (spam scoring up to, but not including this value) 0 %

□ Enable spam score threshold 2
Threshold 2 score (from threshold 1 up to, but not including this value) 95 %

☑ Enable spam score threshold 3
Threshold 3 score (from threshold 2 up to, but not including this value) 97.0000 %
Apply services on this event □ □ □ □ ☑
□ Flag the message as 'Junk' for client handling
□ Enable subject rewrite
Rewrite subject as

☑ Enable spam score threshold 4
Threshold 4 score (from threshold 3 up to, but not including this value) 99.9000 %
Apply services on this event ☑ □ □ □ ☑
□ Flag the message as 'Junk' for client handling
□ Enable subject rewrite
Rewrite subject as

☑ Enable spam score threshold 5
Threshold 5 score (implicitly anything equal or greater than threshold 4)
Apply services on this event ☑ □ □ □ □
□ Flag the message as 'Junk' for client handling
□ Enable subject rewrite
Rewrite subject as

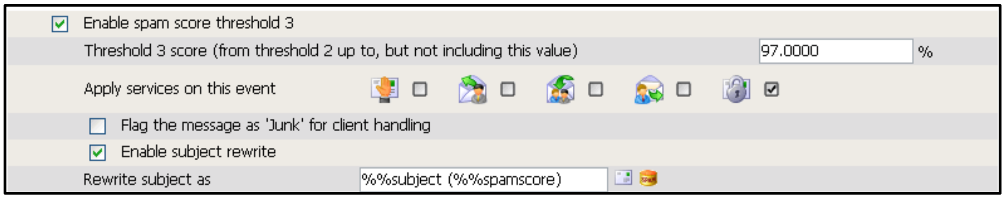
Note: Spam threshold 2 has a percentage of 95 even though that threshold isn't enabled. This is necessary to define a lower limit for spam threshold 3. Spam threshold 3 now refers to messages between 95 and 97 percent.

TIP: In QMS, the administrator will want to configure the digest services to include spam threshold 4 but not spam threshold 3 for this configuration.

Rewriting the Spam Message Subject

GWAVA may be configured to rewrite the subject of an email, depending on the threshold. A common reason to do so is to allow users to create client based rules to treat the messages differently depending on the threshold ("Probably spam" versus "Definitely Spam"). In organizations where individual message filtering is emphasized, it might be appropriate to use this service.

For example, in addition to wanting to place messages with spam probabilities of 95-97% in the quarantine, it might be useful to put the spam percentage in the subject. Check Enable subject rewrite for spam threshold 3 and type in the desired message to include in the subject. The two icons on the right insert variables. The button will insert a variable that will include the original subject. The button will insert a variable that will include the email spam score into the subject line. For example we will use “%%subject (%%spamscore)”, which will return the original subject and then the spam percentage afterwards in parenthesis.



☑ Enable spam score threshold 3

Threshold 3 score (from threshold 2 up to, but not including this value) 97.0000 %

Apply services on this event □ □ □ □ ☑

□ Flag the message as 'Junk' for client handling

☑ Enable subject rewrite

Rewrite subject as %%subject (%%spamscore)

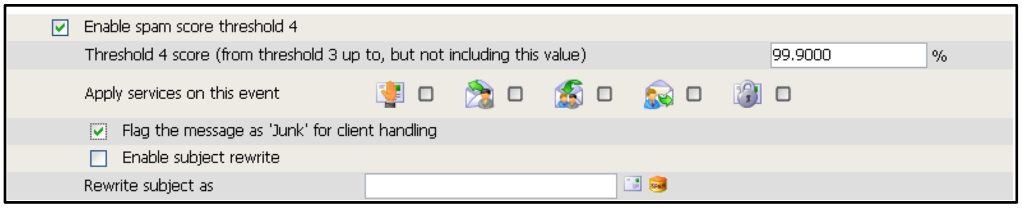
Junk Mail Handling

GWAVA can add “X-Spam-Status: Yes” or “X-Spam-Status: No” to the MIME header of email messages. The email client can then filter on this header field.

Novell GroupWise systems can use this effectively. If the /xspam switch is added to the gwia.cfg and the GWIA is restarted, all messages with “X-Spam-Status: Yes” in the MIME header will automatically be sent to the GroupWise junk mail folder.

Simply check ‘Flag the message as 'Junk' for client handling’ to activate this option.

Keeping with example from above, we will change 'spam threshold 4' to add the X-Spam-Status header. This way any message that GWAVA thinks are spam will end up in the client junk mail folder. The block and quarantine services will be turned off.



☑ Enable spam score threshold 4

Threshold 4 score (from threshold 3 up to, but not including this value) 99.9000 %

Apply services on this event □ □ □ □ □

☑ Flag the message as 'Junk' for client handling

□ Enable subject rewrite

Rewrite subject as

Message Services

A brief explanation of what each service does is given here.

Block - Do not allow this message to be delivered to the user.

Notify Administrator - Send an e-mail to the administrator about the message. The specific contents of the e-mail are based upon a template configured in ‘Configuring Notification Options’

Notify Originator - Send an e-mail to the sender of the message. The specific contents of the e-mail are based upon a template configured in ‘Configuring Notification Options’

Notify Recipient - Send an e-mail to recipient of the message. (Often this only makes sense in conjunction with a Block). The specific contents of the e-mail are based upon a template configured in ‘Configuring Notification Options’

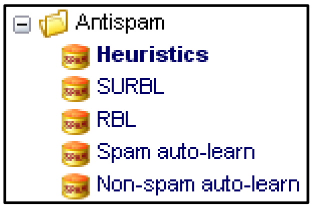
Quarantine - Store the message in the GWAVA Quarantine Management System (QMS). Users with accounts in QMS will be able to access these messages, and possibly release or forward them, depending on their configured rights. These messages also can optionally be included in a HTML digest, if the QMS administrator has turned this function on (Digest Services).

In the screenshot the block service and quarantine service are enabled.

APPENDIX C: SURBL/RBL

SURBL

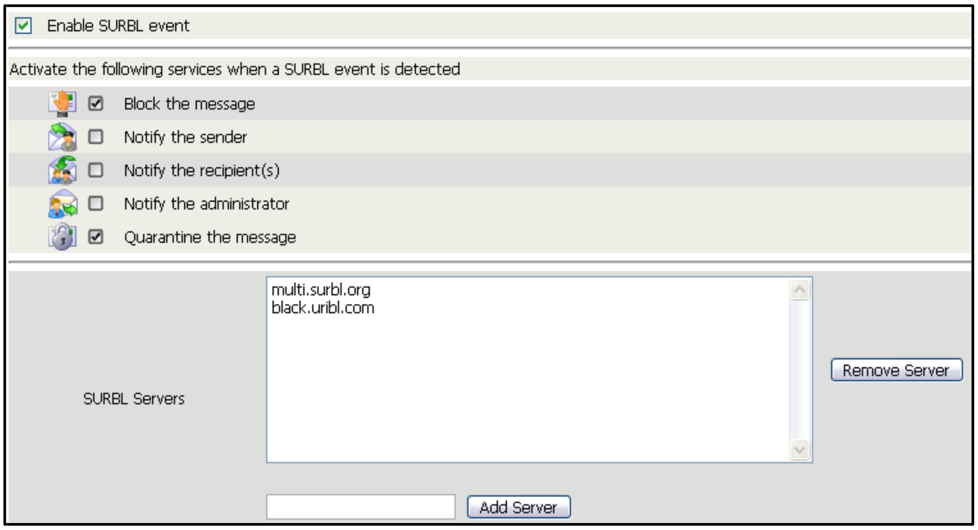
SURBL is a powerful anti-spam technique -- one of the most effective available today. When SURBL is enabled, GWAVA checks the blacklist server(s) to see if the message received contains URLs or links that reference a domain of known spammers.



Check ‘Enable SURBL event’ and select the desired services. To add an additional SURBL server, enter the server hostname in the field provided and click ‘Add Server’.

Default SURBL servers installed by the "Stop Spam" wizard are:

- multi.surbl.org
- black.uribl.com



☑ Enable SURBL event

Activate the following services when a SURBL event is detected

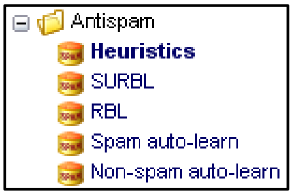
☑ Block the message
□ Notify the sender
□ Notify the recipient(s)
□ Notify the administrator
☑ Quarantine the message

SURBL Servers

multi.surbl.org
black.uribl.com

RBL

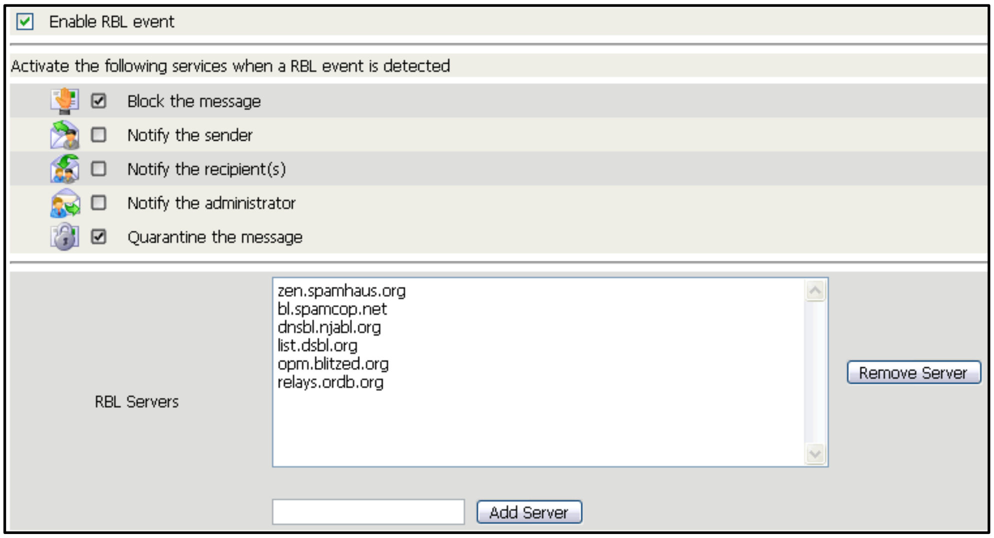
RBL is a common anti-spam technique. With RBL, GWAVA checks the blacklist server(s) to see if the message received is from an IP Address of known spammers. Browse to the RBL section of the 'Antispam' configuration folder.



Check Enable RBL event and select the services you wish to associate with the RBL event. To add an additional RBL server, enter the server hostname in the field provided and click 'Add Server'.

Default RBL servers installed by the "Stop Spam" wizard are:

- zen.spamhaus.org
- bl.spamcop.net
- dnsbl.njabl.org
- list.dsbl.org
- opm.blitzed.org
- relays.ordb.org



☑ Enable RBL event

Activate the following services when a RBL event is detected

☑ Block the message
☐ Notify the sender
☐ Notify the recipient(s)
☐ Notify the administrator
☑ Quarantine the message

RBL Servers

zen.spamhaus.org
bl.spamcop.net
dnsbl.njabl.org
list.dsbl.org
opm.blitzed.org
relays.ordb.org