

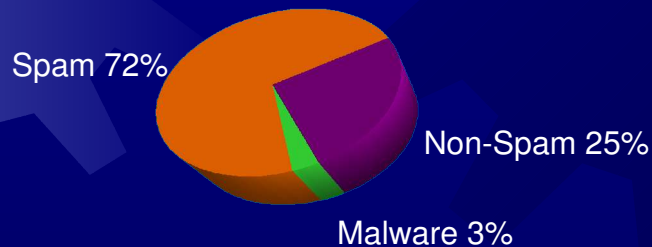


# Open Source Tools for Spam and Virus Management

Robert LeBlanc, Renaissance, Inc.  
[rjl@renaissance.com](mailto:rjl@renaissance.com)

# Scary Numbers

Current Statistics for  
rjl@renaissoft.com



- 70-80% of all e-mail is spam (Verisign, ePrivacy Group, 2004)
- Spam volume is increasing 18% per month (ePrivacy Group, 2003)
- At its peak, 1 in 12 e-mails were MyDoom viruses (MessageLabs, 2004)
- 1,200 new viruses each month (Sophos, 2002)
- Worldwide cost of viruses in 2003: \$55 billion (Trend Micro, 2004)
- Worldwide cost of spam in 2003: \$20.5 billion (Radicati Group, 2003)
- Projected cost of spam in 2007: \$198 billion (Radicati Group, 2003)
- Average cost of spam per employee: \$1,934/year (Nucleus Research, 2004)

# Essential Tasks

- Detect Viruses
- Recognize Spam
- Block and/or Quarantine Mail
- Manage the Process

# Detecting Viruses

Clam AntiVirus <http://www.clamav.net/>

- Runs in command-line and daemonized modes
- On-access scanning under Linux
- 33,000+ viruses, worms, and Trojans recognized
- Virus samples are submitted by the user community
- Ranked 5<sup>th</sup> out of 24 virus scanners tested for responsiveness to new outbreaks (McAfee: 18<sup>th</sup>, Norton: 19<sup>th</sup>) (AV-Test.org, 2004)



# Detecting Viruses

- Use an array of virus scanners from different vendors (in parallel or in series) for maximum protection
- Use daemonized virus scanners whenever possible
- Keep virus signatures up-to-date
- Submit signatures/samples of new viruses

# Recognizing Spam

SpamAssassin <http://spamassassin.apache.org/>

- Pattern-based feature recognizer
- DNSBL, RHSBL, and SPF support
- Collaborative networks: Razor, Pyzor, DCC
- Bayesian classifier
- Comprehensive scoring approach
- Usable at the mail server or on the desktop
- Easily customized and extensible with plug-ins
- New patterns contributed by the user community

# Recognizing Spam

## Feature Recognition

- What is a “feature” of spam?
  - ♦ Spamware signatures
  - ♦ Header inconsistencies
  - ♦ HTML body containing only an image
  - ♦ Obfuscated text
  - ♦ Common scam phrases (e.g. Nigerian Letter formulas)
  - ♦ Etc.
- Use regular expressions to express patterns
- Combinations of patterns can be used: “meta”-rules

# Recognizing Spam

## DNS-Based Tests

- DNSBLs (Domain Name Service Block Lists)
  - ◆ Tests the peer's IP address against various blacklists
  - ◆ Assigns configurable scores to each DNSBL sub-result
- RHSBLs (Right-Hand-Side Block Lists)
  - ◆ Tests URLs in the body of the mail against blacklists
  - ◆ Each RHSBL can be scored independently
- SPF (Sender Policy Framework)
  - ◆ Determines whether the peer is an authorized mail sender for the claimed domain

# Recognizing Spam

## Collaborative Reporting Networks

- Millions of people receive (virtually) identical spam
- 100,000 lemmings can't be wrong!
- Databases store “fuzzy” checksums of reported spam
  - ♦ Vipul's Razor <http://razor.sourceforge.net/>
  - ♦ Pyzor <http://pyzor.sourceforge.net/>
  - ♦ The Distributed Checksum Clearinghouse  
<http://www.rhyolite.com/anti-spam/dcc/>
- SpamAssassin assigns configurable scores to each of these reporting networks

# Recognizing Spam

## Bayesian Classifier

- An “automated feature recognizer”
- Analyzes the frequency with which certain tokens appear in confirmed spam vs. confirmed ham
- Returns a “confidence level” that the mail is spam
- SpamAssassin assigns scores to confidence ranges
- Requires regular “training”

# Recognizing Spam

## SpamAssassin Scoring Example

<u>Score</u>	<u>Rule Name</u>	<u>Rule Description</u>
3.511	PYZOR_CHECK	Listed in Pyzor ( <a href="http://pyzor.sourceforge.net/">http://pyzor.sourceforge.net/</a> )
2.101	BAYES_90	Bayesian spam probability is 90 to 99%
1.113	RCVD_IN_SBL	Received via a relay in the Spamhaus Block List
1.047	RAZOR2_CHECK	Listed in Razor2 ( <a href="http://razor.sourceforge.net/">http://razor.sourceforge.net/</a> )
0.876	RAZOR2_CF_RANGE_11_50	Razor2 gives confidence between 11 and 50%
0.705	MSGID_FROM_MTA_HEADER	Message-ID was added by a relay
0.336	HTML_WEB_BUGS	Image tag intended to identify you
0.320	MIME_HTML_ONLY	Message only has text/html part MIME parts
0.100	HTML_MESSAGE	HTML included in message
0.100	SPAMCOP_URI_RBL	URI's domain appears in sc.surbl.org
<b>10.209</b>		

Each rule contributes “evidence” toward the final score, which is judged against the recipient’s threshold value (5.0 is typical).

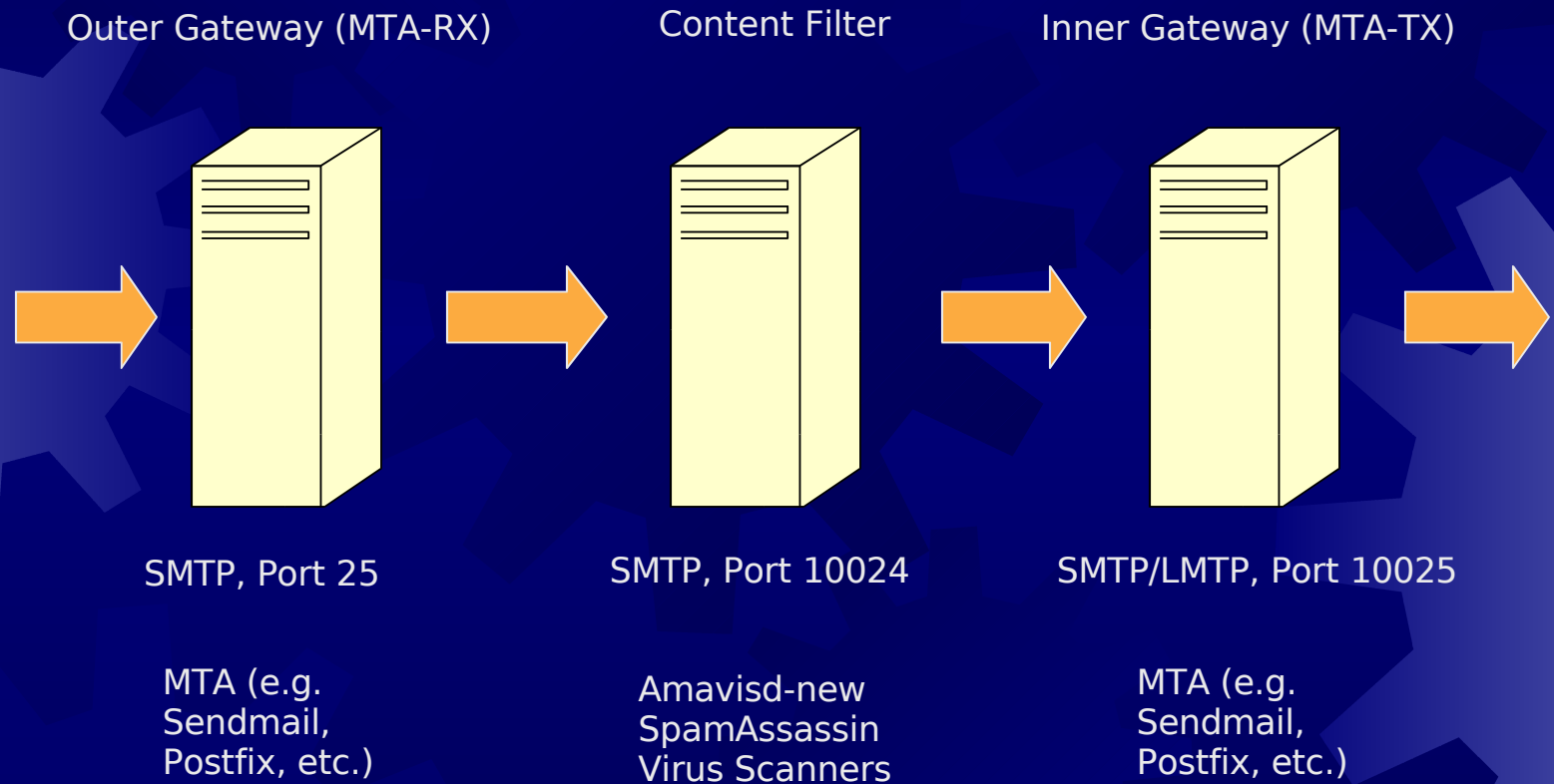
# Blocking and/or Quarantining Suspicious Mail

amavisd-new <http://www.ijs.si/software/amavisd/>

- Content-filtering framework for virus scanners and spam detectors
- Identifies dangerous file attachments
- Objectionable items can be rejected, discarded, quarantined, or passed on
- Per-user, per-domain, and system-wide whitelists and blacklists
- Per-user spam score thresholds
- Per-user overrides for viruses, spam, and attachments



# Blocking and/or Quarantining Suspicious Mail

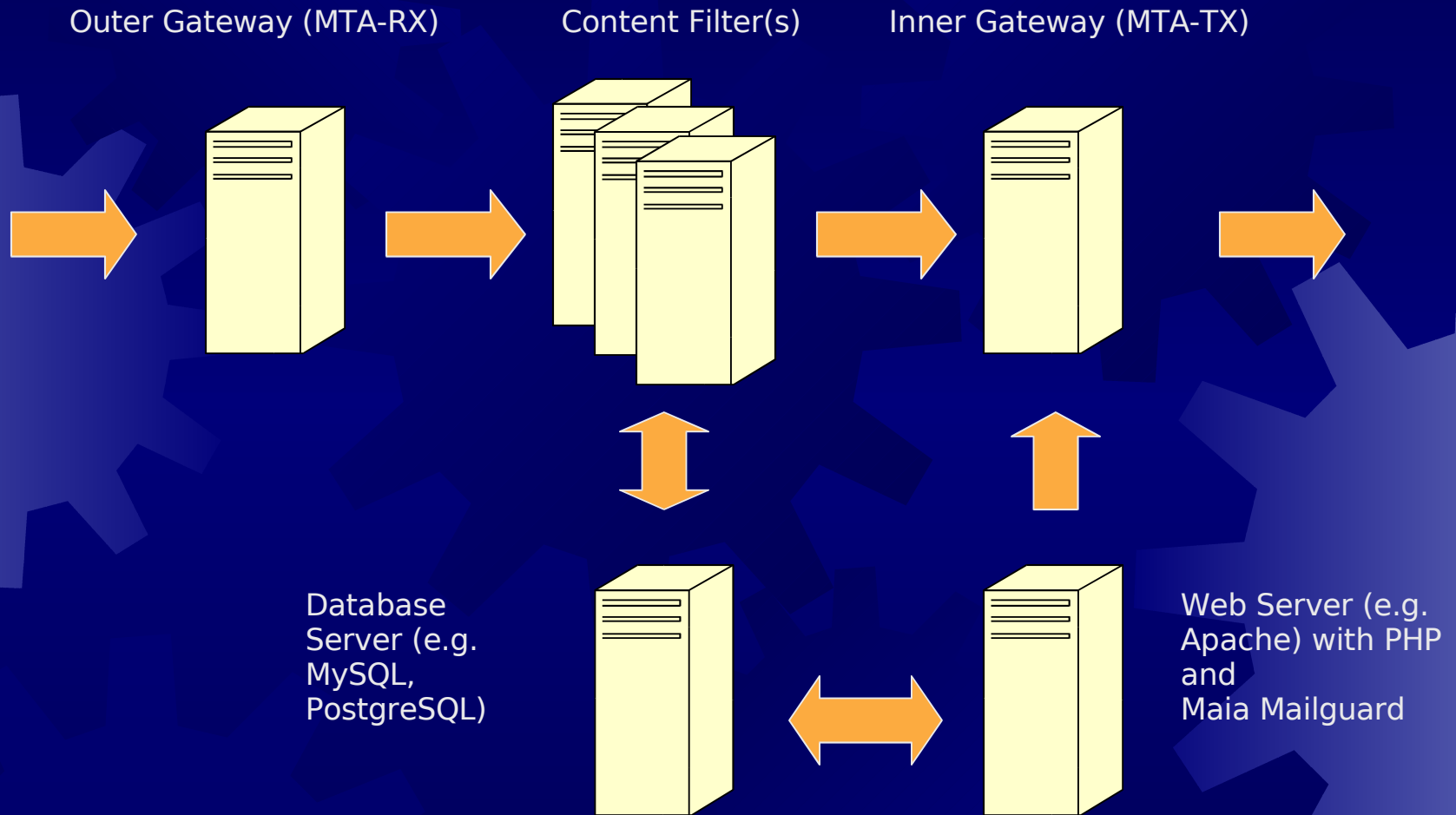


# Managing the Process

Maia Mailguard <http://www.maiamailguard.com/>

- Web-based content filter management
- Users manage their own content filter settings
- Users manage their own whitelists and blacklists
- Users release their own quarantined items
- Users report the spam that gets through
- Users confirm the status of mail as ham or spam
- Confirmed spam and ham is used to train the Bayes database, and spam is reported to collaborative networks
- Manage arrays of content filters from one interface

# Managing the Process



# Managing the Process

## Current protection level: Custom





- ☐ Off
- ☐ Low
- ☐ Medium
- ☒ High

\*\*Custom levels are in use:

Use settings screen to manage, or choose a preset level above.

[Change Level](#)

## Cache Contents

 [Report/Confirm]	You have <b>18</b> items in your ham cache. Click <a href="#">here</a> to help train the filter, or to report a spam message that was missed.
 [Report/Rescue]	You have <b>84</b> items in your spam cache. Click <a href="#">here</a> to report it, or to rescue a message that was mistakenly blocked.
 [Delete/Rescue]	You have <b>3</b> items in your virus cache. Click <a href="#">here</a> to delete it, or to rescue a message that was mistakenly blocked.
 [Delete/Rescue]	You have <b>0</b> items in your banned-file cache. Click <a href="#">here</a> to delete it, or to rescue a message that was mistakenly blocked.
 [Delete/Rescue]	You have <b>0</b> items in your bad-header cache. Click <a href="#">here</a> to delete it, or to rescue a message that was mistakenly blocked.
<a href="#">Delete all items</a>	

**85168** SPAM items have been blocked for you **4289** Viruses have been blocked for you

**200938** Spam items blocked systemwide **12568** Viruses blocked systemwide

# Managing the Process

Statistics for All Users											
Items				Score			Size (kB)			Bandwidth/day	
Mail Type	Count	Items/day	Pct	Min	Max	Avg	Min	Max	Avg	MB	Cost (\$CDN)
Suspected Ham	8357	128.3	1.7%	-20.523	4.981	-1.123	0.6	4412.0	19.5	2.45	0.018
Confirmed Ham	32596	66.2	6.8%	-21.893	21.904	-1.075	0.3	4882.4	12.1	0.78	0.006
False Positives	551	1.1	0.1%	3.692	108.147	7.801	0.6	213.6	20.8	0.02	0.000
Suspected Spam	26261	273.5	5.5%	0.000	140.542	27.685	0.3	207.4	5.6	1.49	0.011
Confirmed Spam	200938	408.4	41.7%	0.000	232.153	34.001	0.3	254.9	4.1	1.65	0.012
False Negatives	707	1.5	0.1%	-16.032	73.710	4.478	0.4	1259.2	19.7	0.03	0.000
Whitelisted Items	196947	407.7	40.9%	-	-	-	0.3	4778.4	7.9	3.15	0.023
Blacklisted Items	54	0.4	0.0%	-	-	-	0.7	25.0	10.2	0.00	0.000
Viruses/Malware	12568	26.4	2.6%	-	-	-	0.7	2299.2	44.8	1.15	0.008
Banned Attachments	2290	5.0	0.5%	-	-	-	0.8	682.0	33.5	0.16	0.001
Invalid Mail Headers	80	0.2	0.0%	-	-	-	0.3	1507.8	55.9	0.01	0.000
Oversized Items	30	0.1	0.0%	-	-	-	5018.7	23338.2	6786.0	0.42	0.003
Efficiency 99.46% False Positive 0.23% False Negative 0.30%											
Sensitivity 99.65% PPV 99.73% Specificity 98.34% NPV 97.88%											



# Managing the Process

Address: rjl@renaissoft.com	
Virus Scanning	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled
Detected viruses should be...	<input type="radio"/> Labeled <input checked="" type="radio"/> Quarantined <input type="radio"/> Discarded
Spam Filtering	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled
Detected spam should be...	<input type="radio"/> Labeled <input checked="" type="radio"/> Quarantined <input type="radio"/> Discarded
Add a prefix to the subjects of spam?	<input type="radio"/> Yes <input checked="" type="radio"/> No
Add X-Spam: Headers when Score is >=	<input type="text" value="-999.000"/>
Consider mail 'Spam' when Score is >=	<input type="text" value="5.000"/>
Quarantine Spam when Score is >=	<input type="text" value="5.000"/>
Attachment Type Filtering	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled
Mail with dangerous attachments should be...	<input type="radio"/> Labeled <input checked="" type="radio"/> Quarantined <input type="radio"/> Discarded
Bad Header Filtering	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled
Mail with bad headers should be...	<input type="radio"/> Labeled <input checked="" type="radio"/> Quarantined <input type="radio"/> Discarded

# Managing the Process

Suspected Spam ( 1 - 86 of 86 )						
Confirm the Status of these Items						
Score	Received	From	Subject	<input type="radio"/> Spam?	<input type="radio"/> Ham?	<input type="radio"/> Delete
6.6	2005-04-28 14:11:20	rjl@renaissoft.com	Here is your trac...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9.4	2005-04-29 00:08:15	steveaohl@stevefe...	Innocent russian ...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10.5	2005-04-28 21:42:14	cd2004@compaqnet.fr	Telephone sans fi...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10.6	2005-04-28 19:22:52	fuqscaocus@freema...	Online money	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11.9	2005-04-28 13:28:28	tpuoj@mlmcentral.com	Re: topple Levami...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12.7	2005-04-28 13:50:16	henriksenbaoqin@a...	what happend to cex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12.9	2005-04-28 23:17:09	rqqemvbhepxdj@con...	Hot st0ck tips fr...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12.9	2005-04-28 20:45:12	vecdkmk@fula.co.uk	This company is c...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13.6	2005-04-28 15:56:54	rvjeng@jhurry.fre...	Offering Funding ...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13.7	2005-04-28 17:07:52	swkiakt@contact-o...	st0ck rum0rs that...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13.9	2005-04-29 01:04:11	gnkxkybcdevqy@cat...	Better st0ck perf...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13.9	2005-04-29 00:39:53	polyzvnwcxcid@div...	US h0t st0ck high...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13.9	2005-04-29 00:33:31	oyqzsakf@dentelas...	Unbiased info for...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14.5	2005-04-28 19:10:05	prohibitivegibson...	Meet easy girls	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14.9	2005-04-28 16:36:12	rnmrg@foxdigital.ch	Acquire at the bo...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

# Managing the Process

## Suspected Ham ( 1 - 18 of 18 )

Confirm the Status of these Items

Score	Received	From	Subject	<input type="radio"/> Spam?	<input type="radio"/> Ham?	<input type="radio"/> Delete
4.8	2005-04-28 15:52:29	rjl@sourcreambblas...	One hand clapping	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-0.3	2005-04-28 14:42:30	wilma2002@spray.se	Re: Re: [Maia-use...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-1.9	2005-04-28 18:25:17	rosander@owbn.org	RE: [Maia-users] ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.4	2005-04-28 19:30:54	francus@metsny.yo...	Re: [Maia-users] ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.5	2005-04-28 22:05:14	fest-list-bounces...	[Fest-list] Bandw...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.5	2005-04-28 13:26:43	chris.paul@sentin...	Re: [Maia-users] ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.5	2005-04-28 14:11:19	chris.paul@sentin...	Re: [Maia-users] ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.5	2005-04-28 18:27:28	segr@segr.ca	Re: [Maia-users] ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 23:12:22	fest-list-bounces...	Re: [Fest-list] N...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 19:05:40	fest-list-bounces...	[Fest-list] I'm h...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 20:49:47	fest-list-bounces...	Re: [Fest-list] I...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 22:30:52	fest-list-bounces...	[Fest-list] Prese...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 22:56:27	fest-list-bounces...	Re: [Fest-list] I...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-2.6	2005-04-28 23:00:55	fest-list-bounces...	[Fest-list] Not r...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-3.7	2005-04-29 00:06:48	googlealerts-nore...	Google Alert - am...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-12.8	2005-04-29 01:24:55	emailsenderapp+19...	Unique Mother's D...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
-20.4	2005-04-28 12:45:24	payment@paypal.com	Receipt for your ...	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>



# Managing the Process

Score	Rule Triggered	Explanation
3.511	PYZOR_CHECK	Listed in <a href="http://pyzor.sf.net/">Pyzor</a> ( <a href="http://pyzor.sf.net/">http://pyzor.sf.net/</a> )
2.599	RCVD_IN_DYNABLOCK	Sent directly from dynamic IP address
1.657	BAYES_80	Bayesian spam probability is 80 to 90%
1.500	RCVD_IN_BL_SPAMCOP_NET	Received via a relay in <a href="http://bl.spamcop.net">bl.spamcop.net</a>
1.000	BAYES_POISON_25	25-49 unpunctuated lower-case words in a row
0.706	RCVD_IN_DSBL	Received via a relay in <a href="http://list.dsbl.org">list.dsbl.org</a>
0.100	RCVD_IN_SORBS	<a href="http://www.sorbs.net">SORBS</a> : sender is listed in <a href="http://www.sorbs.net">SORBS</a>
0.100	HTML_MESSAGE	HTML included in message


**FROM:** "Spammer" <spammer@example.com>  
**TO:** jsmith@example.com  
**SUBJECT:** Make home improvements, or take-cash out  
**CONTENT-TYPE:** multipart/alternative

**CONTENT-TYPE:** text/plain

planeload melanin zeal soul manslaughter kerr once fourteenth cagey ow  
perfume ambiguity breeches boisterous junction applicant baghdad  
tollhouse vitamin deactivate pitchfork continual anecdote christlike  
moulton advantageous select minnesota allocable imperate cominform  
marrow gotham bug savonarola bonn

**CONTENT-TYPE:** text/html

Re-finance now, even with bad-credit!

 \*Best Re-finance Rate for credit challenged.  
\*Best Customer Service  
\*Lowest Interest-Rates in Years  
\*SAVE \$100-\$400 per month

Our easy application only takes 1 minutes.

[Visit here for more information](#)

# Managing the Process

E-mail address or domain to add:

List to add to: ☒ Whitelist ☐ Blacklist

Address	Whitelist	Blacklist	Remove
akhan@example.com	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
ldavis@example.com	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
tmoore@example.com	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<input type="button" value="Update"/>		<input type="button" value="Reset"/>	

# Managing the Process

System Configuration	
Enable auto-creation of user accounts? [?]	<input type="radio"/> Yes <input checked="" type="radio"/> No
Enable tracking of false negatives? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable tracking of statistics? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable virus scanning? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable spam filtering? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable banned file attachment checks? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable checks for invalid mail headers? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Enable spam-trap accounts? [?]	<input checked="" type="radio"/> Yes <input type="radio"/> No
Mail size limit (bytes): [?]	<input type="text" value="5000000"/>
Oversized items should be... [?]	<input type="radio"/> Accepted <input checked="" type="radio"/> Rejected

Paths & Ports	
Maia installation directory: [?]	<input type="text" value="/usr/local/apache/htdocs/mail"/>
Administrator's contact e-mail address: [?]	<input type="text" value="jsmith@example.com"/>
Downstream SMTP server (MTA-TX): [?]	<input type="text" value="localhost"/>
Downstream SMTP port number: [?]	<input type="text" value="10025"/>
Encryption key file (optional): [?]	<input type="text" value="/var/amavisd/blowfish.key"/>

Cache Expiry & Quarantine Reminders	
Expiry period for quarantined mail (days): [?]	<input type="text" value="10"/>
Expiry period for cached ham (days): [?]	<input type="text" value="5"/>
E-mail reminder threshold (items): [?]	<input type="text" value="100"/>
E-mail reminder threshold (size): [?]	<input type="text" value="500000"/>
E-mail reminder template file: [?]	<input type="text" value="/var/amavisd/maia/templates/reminder.tpl"/>
Maia login URL for e-mail reminders: [?]	<input type="text" value="http://www.example.com/mail/"/>

# Managing the Process



**Maia Mailguard**  
**Welcome**

User: rjl

Cache Contents:

Suspected Ham: 132

Suspected Spam: 89

Virus/Malware: 3



[ Welcome ]



[ Stats ]



[ W/B List ]



## Maia Mailguard

### Mail Filter Settings

User: rjl



#### Statistics for All Users

Mail Type	Items			Score			Size (kB)			Bandwidth/day	
	Count	Items/day	Pct	Min	Max	Avg	Min	Max	Avg	MB	Cost (\$CDN)
Suspected Ham	8362	128.4	1.7%	-20.523	4.981	-1.126	0.6	4412.0	19.5	2.45	0.018
Confirmed Ham	32596	66.2	6.8%	-21.893	21.904	-1.075	0.3	4882.4	12.1	0.78	0.006
False Positives	551	1.1	0.1%	3.692	108.147	7.801	0.6	213.6	20.8	0.02	0.000
Suspected Spam	26299	273.8	5.5%	0.000	140.542	27.685	0.3	207.4	5.6	1.49	0.011
Confirmed Spam	200938	408.4	41.7%	0.000	232.153	34.001	0.3	254.9	4.1	1.65	0.012
False Negatives	707	1.5	0.1%	-16.032	73.710	4.478	0.4	1259.2	19.7	0.03	0.000
Whitelisted Items	196947	407.7	40.9%	-	-	-	0.3	4778.4	7.9	3.15	0.023
Blacklisted Items	54	0.4	0.0%	-	-	-	0.7	25.0	10.2	0.00	0.000
Viruses/Malware	12568	26.4	2.6%	-	-	-	0.7	2299.2	44.8	1.15	0.008
Banned Attachments	2290	5.0	0.5%	-	-	-	0.8	682.0	33.5	0.16	0.001
Invalid Mail Headers	80	0.2	0.0%	-	-	-	0.3	1507.8	55.9	0.01	0.000
Oversized Items	30	0.1	0.0%	-	-	-	5018.7	23338.2	6786.0	0.42	0.003

Efficiency 99.46% False Positive 0.23% False Negative 0.30%  
Sensitivity 99.65% PPV 99.73% Specificity 98.34% NPV 97.88%

# On the Horizon

- Dynamic DNSBL support
- Greylisting/tarpitting support
- Graphical charts
- The Maia Network
- More reporting mechanisms
- Maia Mailguard Appliances



# Open Source Rules!

You don't need to spend a fortune to get world-class spam and virus management tools—open source tools lead the way:

- Clam AntiVirus
- SpamAssassin
- Vipul's Razor, Pyzor, and DCC
- amavisd-new
- Maia Mailguard
- Perl, PHP, Apache, MySQL, PostgreSQL, Sendmail, Postfix...
- ...and of course ***Linux!***