

iRedMail/FAQ/Store.SpamAssassin.Bayes.In.SQL

From iRedMail

Contents

- 1 Summary
- 2 Create required SQL database used to store bayes data
- 3 Enable Bayes modules in SpamAssassin
- 4 Enable Roundcube plugin: markasjunk2
- 5 References

THIS ARTICLE IS STILL A DRAFT, DO NOT APPLY IT IN PRODUCTION SERVER.

Summary

This article is used to configure related components to store SpamAssassin Bayes data in SQL server, and allow webmail users to report spam with one click.

Tested with:

- iRedMail-0.8.0 with MySQL backend.
- CentOS 6.2 (x86_64)
- SpamAssassin-3.3.1
- Amavisd-new-2.6.6
- MySQL-5.1.61
- Roundcubemail-0.7.2

Notes:

- This article should work with all iRedMail releases. We take iRedMail-**0.8.0** for example.
- This article should work with all backends: OpenLDAP, MySQL, PostgreSQL. We take MySQL backend for example.
- This article should work with Amavisd-new-2.6.0 and later versions, includes Amavisd-new-2.7.x.

IMPORTANT NOTE:

- The bayesian classifier can only score new messages if it already has 200 known spams and 200 known hams.
- If Spamassassin fails to identify a spam, teach it so it can do better next time. e.g. Mark it as spam in roundcube webmail.
- Read **References** section at the end of this article before asking.

Create required SQL database used to store bayes data

We need to create a SQL database and necessary tables to store SpamAssassin bayes data. The RPM package installed on CentOS 6 doesn't ship SQL template for bayes database, but we can download it from Apache web site. We're running SpamAssassin-3.3.1, so what we need is this SQL template file: http://svn.apache.org/repos/asf/spamassassin/tags/spamassassin_release_3_3_1/sql/bayes_mysql.sql (if you're running different version, please find the proper SQL file here: <http://svn.apache.org/repos/asf/spamassassin/tags/>)

Terminal:

```
# cd /root/
# wget http://svn.apache.org/repos/asf/spamassassin/tags/spamassassin_release_3_3_1/sql/bayes_mysql.sql
```

Create MySQL database and import SQL template file:

Terminal:

```
# mysql -uroot -p
mysql> CREATE DATABASE sa_bayes;
mysql> USE sa_bayes;
mysql> SOURCE /root/bayes_mysql.sql;
```

Create a new MySQL user (with password **sa_user_password**) and grant permissions:

- Note: Please replace password **sa_user_password** by your own.

Terminal:

```
mysql> GRANT SELECT, INSERT, UPDATE, DELETE ON sa_bayes.* TO sa_user@localhost IDENTIFIED BY 'sa_user_password';
mysql> FLUSH PRIVILEGES;
```

Enable Bayes modules in SpamAssassin

Edit /etc/mail/spamassassin/local.cf, add (or modify below settings):

File: *local.cf*

```
#use_bayes 1
#bayes_auto_learn 1
#bayes_auto_expire 1

# Store bayesian data in MySQL
#bayes_store_module Mail::SpamAssassin::BayesStore::MySQL
#bayes_sql_dsn DBI:mysql:sa_bayes:127.0.0.1:3306

# Store bayesian data in PostgreSQL
#bayes_store_module Mail::SpamAssassin::BayesStore::PgSQL
#bayes_sql_dsn DBI:Pg:sa_bayes:127.0.0.1:5432

#bayes_sql_username sa_user
#bayes_sql_password sa_user_password

# Override the username used for storing
# data in the database. This could be used to group users together to
# share bayesian filter data. You can also use this config option to
# trick sa-learn to learn data as a specific user.
#bayes_sql_override_username vmail
```

Make sure SpamAssassin will load bayes modules:

Terminal:

```
# /etc/init.d/amavisd stop
# amavisd -c /etc/amavisd/amavisd.conf debug 2>&1 | grep -i 'bayes'
May 16 09:59:33 ... SpamAssassin loaded plugins: ..., Bayes, ...
May 16 10:27:38 ... extra modules loaded after daemonizing/chrooting:
    Mail/SpamAssassin/BayesStore/MySQL.pm, Mail/SpamAssassin/BayesStore/SQL.pm, ...
```

Looks fine. Now press 'Ctrl-C' to terminate above command.

Start Amavisd service:

Terminal:

```
# /etc/init.d/amavisd restart
```

It is required we initialize the database by learning a message. We use the sample spam email shipped in the RPM package provided by CentOS 6:

Terminal:

```
# rpm -ql spamassassin | grep 'sample-spam'
/usr/share/doc/spamassassin-3.3.1/sample-spam.txt

# sa-learn --spam --username=vmail /usr/share/doc/spamassassin-3.3.1/sample-spam.txt
Learned tokens from 1 message(s) (1 message(s) examined)
```

Enable Roundcube plugin: markasjunk2

- We need a third-party Roundcube plugin to allow webmail users to report spam: Mark as Junk 2. You can download it here: <http://www.tehinterweb.co.uk/roundcube/#pimarkasjunk2>
- After download, please uncompress it and copy it to roundcube plugins directory: `/var/www/roundcubemail/plugins/`. Then we get a new directory: `/var/www/roundcubemail/plugins/markasjunk2/`
- Enter directory `/var/www/roundcubemail/plugins/markasjunk2/`, generate config file by copying its sample config file:

Terminal:

```
# cd /var/www/roundcubemail/plugins/markasjunk2/
# cp config.inc.php.dist config.inc.php
```

- Edit `config.inc.php`, update below settings:

File: **`roundcubemail/plugins/markasjunk2/config.inc.php`**

```
$rcmail_config['markasjunk2_learning_driver'] = 'cmd_learn';
$rcmail_config['markasjunk2_read_spam'] = true;
$rcmail_config['markasjunk2_unread_ham'] = false;
$rcmail_config['markasjunk2_move_spam'] = true;
$rcmail_config['markasjunk2_move_ham'] = true;
$rcmail_config['markasjunk2_mb_toolbar'] = true;

$rcmail_config['markasjunk2_spam_cmd'] = 'sa-learn --spam --username=vmail %f';
$rcmail_config['markasjunk2_ham_cmd'] = 'sa-learn --ham --username=vmail %f';
```

- Enable this plugin in Roundcube config file by appending 'markasjunk2' in plugin list:

File: **`/var/www/roundcubemail/config/main.inc.php`**

```
$rcmail_config['plugins'] = array("password", "managesieve", "markasjunk2");
```

- Since learning driver `cmd_learn` requires PHP function `exec`, we have to enable it in `/etc/php.ini`:

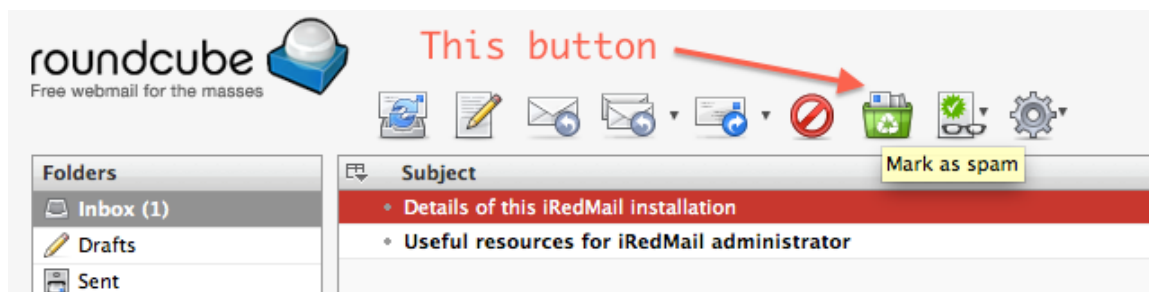
File: **`/etc/php.ini`**

```
# OLD SETTING
# disable_functions =show_source,system,shell_exec,passthru,exec,phpinfo,proc_open ;

# NEW SETTING. exec is removed.
disable_functions =show_source,system,shell_exec,passthru,phpinfo,proc_open ;
```

- Restarting Apache web server.

You will see a new toolbar button after logging into Roundcube webmail:



Check SQL database **sa_bayes** before we testing this plugin:

Terminal:

```
# mysql -uroot -p
mysql> USE sa_bayes;
mysql> SELECT COUNT(*) FROM bayes_token;
+-----+
| count(*) |
+-----+
|        65 |
+-----+
```

Back to Roundcube webmail, select a spam email (or a testing email), click "Mark as Junk" button, then this email will be scanned by command **sa-learn**. Check database **sa_bayes** again to make sure it's working:

Terminal:

```
# mysql -uroot -p
mysql> USE sa_bayes;
mysql> SELECT COUNT(*) FROM bayes_token;
+-----+
| count(*) |
+-----+
|       143 |
+-----+
```

Note: You may get different result number as shown above.

So far so good. That's all we need to do.

References

- Bayes Introduction (<http://wiki.apache.org/spamassassin/BayesInSpamAssassin>) . Please do read section **Things to remember**.
- SpamAssassin Bayes Frequently Asked Questions (<http://wiki.apache.org/spamassassin/BayesFaq>)

Retrieved from "<http://www.iredmail.org/wiki/index.php?title=iRedMail/FAQ/Store.SpamAssassin.Bayes.In.SQL&oldid=2149>"

- This page was last modified on 9 June 2012, at 07:57.