Storing SpamAssassin User Data in SQL Databases
Michael Parker

Introduction

- General Database Setup

- User Preferences

- Auto-Whitelist

- Bayes

- Future Directions

General Database Setup

- Supported Databases

- Determine your Data Source Name (DSN)

- Database Architecture

- Setting up your Database

Determine Your Data Source Name

- DSN varies based on Database Driver

- Examples:

  - dbi:mysql:spamassassin

  - dbi:Pg:dbname=spamassassin

  - dbi:SQLite:dbname=/path/spamassassin.db

Database Architecture Considerations

- Multiple Databases

- Multiple Machines

- Multiple Engines

Basic Database Setup

MySQL

mysql -u <adminusername> -p mysql

Enter password: <adminpassword>

mysql> create database spamassassin;

mysql> grant all on spamassassin.* to sauser@localhost identified by 'spamki11er';

PostgreSQL

psql -U <adminusername> template1

template1=# CREATE USER sauser PASSWORD 'spamki11er';

template1=# CREATE DATABASE spamassassin OWNER = sauser;

User Preferences

- Setup and Configuration

- Populating userpref Table

- GUI Frontends

- Custom SQL Query

User Preferences Database Setup

```
CREATE TABLE userpref (
  username varchar(100) NOT NULL default '',
  preference varchar(30) NOT NULL default '',
  value varchar(100) NOT NULL default '',
  prefid int(11) NOT NULL auto_increment,
  PRIMARY KEY  (prefid),
  KEY username (username)
) TYPE=MyISAM;


mysql -u<adminusername> -p spamassassin < userpref_mysql.sql
Enter password: <adminpassword>
```

User Preferences Configuration

user_scores_dsn   dbi:mysql:spamassassin

user_scores_sql_username   sauser

user_scores_sql_password   spamki11er

spamd Startup and Debug

- Must use the -q or -Q options

- Use -x option to keep from looking up user_prefs

/usr/bin/spamd -d -x -q

logmsg: connection from localhost [127.0.0.1] at port 34601
debug: Conf::SQL: executing SQL: select preference, value ...
debug: retrieving prefs for joe@example.com from SQL server
debug: user has changed

Populating userpref Table

insert into userpref (username, preference, value)
  VALUES ('@GLOBAL', 'required_score', '10.0');

insert into userpref (username, preference, value)
  VALUES ('@GLOBAL', 'whitelist_from', '*@apache.org');

insert into userpref (username, preference, value)
  VALUES ('joe@example.com', 'required_score', '5.5');

insert into userpref (username, preference, value)
  VALUES ('joe@example.com', 'whitelist_from', 'sue@betty.org');

insert into userpref (username, preference, value)
  VALUES ('joe@example.com', 'blacklist_from', '*@spamking.com');

GUI Frontends

http://wiki.apache.org/spamassassin/WebUserInterfaces

Creating a Custom SQL Query

Current default query:
   SELECT preference, value FROM _TABLE_
   WHERE username = _USERNAME_
      OR username = '@GLOBAL' ORDER BY username ASC

 Use global and then domain level defaults:
   SELECT preference, value FROM _TABLE_
   WHERE username = _USERNAME_
      OR username = '@GLOBAL'
      OR username = '@~'||_DOMAIN_ ORDER BY username ASC

Maybe global prefs should override user preferences:
   SELECT preference, value FROM _TABLE_
   WHERE username = _USERNAME_
      OR username = '@GLOBAL' ORDER BY username DESC"

Custom SQL Query Variables

USERNAME
> The current user's username.

MAILBOX
> The portion before the @ as derived from the current username.

DOMAIN
> The portion after the @ as derived from the current username, this value may be null.

Auto-Whitelist

- Database Table Setup

- Configuration

- Expiration

Auto-Whitelist Database Setup

```
CREATE TABLE awl (
  username varchar(100) NOT NULL default '',
  email varchar(200) NOT NULL default '',
  ip varchar(10) NOT NULL default '',
  count int(11) default '0',
  totscore float default '0',
  PRIMARY KEY  (username,email,ip)
) TYPE=MyISAM;



mysql -u<adminusername> -p spamassassin < awl_mysql.sql
Enter password: <adminpassword>
```

Auto-Whitelist Configuration

use_auto_whitelist   1

user_awl_dsn   dbi:mysql:spamassassin
user_awl_sql_username   sauser
user_awl_sql_password   spamki11er

auto_whitelist_factory   Mail::SpamAssassin::SQLBasedAddrList

Your Mother Doesn't Live Here, Clean up Your Own Mess

- Auto-Whitelist has no Expiration

- With SQL you can fake it

```
CREATE TABLE awl (
  username varchar(100) NOT NULL default '',
  email varchar(200) NOT NULL default '',
  ip varchar(10) NOT NULL default '',
  count int(11) default '0',
  totscore float default '0',
  lastupdate timestamp(14) NOT NULL,
  PRIMARY KEY  (username,email,ip)
) TYPE=MyISAM;
```

Roll Your Own Expiration

Delete all entries not updated in the last 6 months:

```
mysql> DELETE FROM awl
        WHERE lastupdate <= DATE_SUB(SYSDATE(), INTERVAL 6 MONTH);
```

Delete all single count entries that haven't been updated in the last 30 days:

```
mysql> DELETE FROM awl WHERE count = 1

        AND lastupdate <= DATE_SUB(SYSDATE(), INTERVAL 30 DAY);
```

Delete possibly statistically insignificant rows:

```
mysql> DELETE FROM awl WHERE totscore/count < .1 AND totscore/count > -.1;
```

Bayes

- Database Setup

- Configuration

- Converting from DBM to SQL

- Bayes with PostgreSQL

Bayes Database Setup

- 6 Different Tables

mysql -u <adminusername> -p spamassassin < bayes_mysql.sql
Enter password: <adminpassword>

Bayes Configuration

bayes_sql_dsn   dbi:mysql:spamassassin

bayes_sql_username   sauser

bayes_sql_password   spamki11er

bayes_store_module   Mail::SpamAssassin::BayesStore::SQL

bayes_sql_override_username   sitewideuser

Converting from DBM to SQL

1) Install 3.0 and follow UPGRADE to bring your databases up to the current version

2) sa-learn --backup > backup.txt

3) Optional: sa-learn --clear to remove old DBM files

4) Switch local.cf to turn on Bayes SQL

5) sa-learn --restore backup.txt

Bayes with PostgreSQL

- Convert token column to BYTEA

- Convert code to use explicit bind parameters

- Move some code to PL/pgSQL

- Transaction Support

Future Directions

- Learning Support for spamd

- Persistent Database Connections

- Better PostgreSQL Support

- Journal Like Support

Where to Get More Info

- Wiki
  - http://wiki.apache.org/spamassassin

- Mailing Lists
  - http://wiki.apache.org/spamassassin/MailingLists

- IRC
  - #spamassassin on irc.freenode.net

- Updated Slides
  - http://www.apache.org/~parker/presentations/index.html