CorpusNightlyScript

Corpus Nightly Script

A script included with the SpamAssassin code for contributing to the NightlyMassChecks. Doesn't seem popular now.

The corpus-nightly script in the masses/rule-qa/ directory of the SpamAssassin tree can be used to set up a mass-checker on your mail. Here's a step-by-step account of the process.

First off, you'll also need to ask for RsyncAccounts and make sure you get a "nightly" account rather than a release-time account. You also need to install Subversion to get the "svn" command.

Then run:

```
mkdir $HOME/nightlymc $HOME/nightlymc/tmp
cd $HOME/nightlymc
svn co http://svn.apache.org/repos/asf/spamassassin/trunk
cp trunk/masses/rule-qa/corpus.example ~/.corpus
```

Edit '~/.corpus' to have values something like this, replacing /home/jm with whatever your own $HOME is.

```
vi ~/.corpus
# temporary working directory for summary results
tmp=/home/jm/nightlymc/tmp

# subversion directory location
# [this is the directory you have already checked out!]
tree=/home/jm/nightlymc/trunk

# rsync username and password (see RsyncAccounts)
username=jm
password=xyzzy

# weekly and nightly mass-check options
opts_weekly="--restart=500 --tail=15000 --net -j 8 -f /home/jm/nightlymc/targets"
opts_nightly="--restart=500 --tail=15000 -f /home/jm/nightlymc/targets"

# weekly and nightly mass-check user_prefs files
prefs_weekly=/home/jm/nightlymc/user_prefs.weekly
prefs_nightly=/home/jm/nightlymc/user_prefs.nightly
```

Now, create those two user_prefs files. Here's suggested (basic) settings:

user_prefs.nightly:

```
use_bayes 0
use_auto_whitelist 0
internal_networks 127/8
trusted_networks 127/8
```

I suggest just "cp"'ing that file to `user_prefs.weekly` as well, but if you wanted different settings to control network rules, go ahead. It might make sense to extend those with full trusted-networks data, if you like.

Edit `~/nightlymc/targets`:

```
ham:detect:/local/cor/recent/ham/*
spam:detect:/local/cor/recent/spam/*
```

That's it – now run
`bash /home/jm/nightlymc/trunk/masses/rule-qa/corpus-nightly` and watch as it starts mass-checking. Once you're happy enough with it, set that command to run in cron.

Note: the best time to run a mass-check is as soon as possible after 0900 UTC. Daylight savings time in some local timezones can be troublesome, so the script will adjust for this by sleeping for an hour if it detects that it was started in the 0800 UTC hour period, so you no longer have to worry about that.