

NightlyMassCheck

Nightly Mass-Check Runs

What?

Nightly [MassCheck](#) runs are the way people submit data on the effectiveness of current rules on their recent spam and ham. It is used to generate the very rule scores that determine the effectiveness of [SpamAssassin](#) (distributed via sa-update), and to evaluate rules via the [RuleQaApp](#). The accuracy of [SpamAssassin](#) is directly related to the number of people contributing to nightly [MassChecks](#).

This does not require sending us your email. Just logs of which rules hit your emails.

Usually a script is run from cron which automatically downloads the latest development version of [SpamAssassin](#), runs it against your spam and ham, and then uploads a log of the results. One line per email, with a list of the [SpamAssassin](#) rules each email hit. **Your actual email is not uploaded with this method.**

An advantage to participating is that it makes [SpamAssassin](#) more accurate on your email. Even few hundred varied emails per month would be good help for the project. But please make sure you are committed to maintaining a [clean corpus](#) and willing to monitor RuleQA Mailing List for queries and updates to masscheck participants.

How?

1. Send an email to private@spamassassin.apache.org requesting an [rsync account](#) for nightly mass-checks. It is helpful if you include a few sentences with your background and expertise for vetting purposes. NOTE: New masscheck contributors are now being accepted since about 2012-08-09.
2. When your request is processed, you'll be notified and added to the [RuleQA Mailing List](#) for [SpamAssassin](#).
3. Download [automasscheck-minimal.sh](#) and [automasscheck-minimal.cf.dist](#)
4. *Local [SpamAssassin](#) installation is not needed nor will be used in any way*, the script downloads a private one for it's own use. But make sure you have [required perl modules](#) installed.
5. Copy `automasscheck-minimal.sh` to `~/bin/` or other suitable location.
6. Copy `automasscheck-minimal.cf.dist` to `~/automasscheck.cf` (hardcoded location in script).
7. Modify `~/automasscheck.cf` to point at your ham and spam folders. Be sure to configure properly for mbox (mbox) or Maildir (dir) folder formats. Leave the RSYNC options unchanged for now, because you will be running automasscheck in test mode at first. Set WORKDIR to suitable location.
8. The masschecker is set to run 8 threads by default. Depending on your box's resources this could too much or too little. Modify your `~/automasscheck.cf` to change JOBS as needed.
9. Set [TRUSTED_NETWORKS](#) and [INTERNAL_NETWORKS](#) properly in `~/automasscheck.cf`. Without this network tests might use the wrong relay.
10. Ensure there is no router/firewall blocking connections to `rsync.spamassassin.org` port 873 (rsync protocol).
11. Run `automasscheck-minimal.sh`.
 - Look in `~/masscheckwork/nightly_mass_check/masses/` for `ham-*.log` and `spam-*.log` files. (Or `weekly_mass_check` on Saturday.)
 - Are the filenames good? They should be named something like `ham-username.log` or `ham-net-username.log`.
 - Read [CorpusCleaning](#) and [HandClassifiedCorpora](#) for guidelines of how to identify ham in your spam folder, and spam in your ham folder, and which messages you should be simply deleted.
 - If you move/delete messages, do not forget to "Compact Folder" to be sure they are actually gone.
 - Repeat automasscheck until you are certain both folders are cleaned.
12. Edit `~/automasscheck.cf` and set RSYNC_USERNAME and RSYNC_PASSWORD with values from step 1.
13. Run `automasscheck-minimal.sh`, which will upload your results.
14. Ask a more experienced participant (probably the person who recruited you) to check your results on the server. They can see the uploaded log files by running a command like `rsync --old-d username@rsync.spamassassin.org::corpus/`. You can also verify that your corpora show up on <http://ruleqa.spamassassin.org/> - the green box near the top shows all included usernames.
15. If your upload looks good, then you're probably ready to automate nightly checks. Configure automasscheck to run as a cron job as your non-root user at or after 9AM UTC. (After `weekly-versions.txt` / `nightly-versions.txt` gets updated in `rsync.spamassassin.org::corpus` . If you run it earlier it will [break things](#) .)

Alternative Methods

You can do it manually: [ManualNightlyMassCheck](#) (but you really should not, just for reference)