

MultiFileRuleSets

Disclaimer

Custom or third-party rules described here are not part of the official [SpamAssassin](#) distribution. They may have a different license and are not from the Apache Software Foundation.

Multi-File Rule Sets

In mid-2004, Bob Menschel, SARE Ninja, started developing multi-file rule sets. This allows us to segregate SARE rules not only by type or subject matter, but also by whether it's applicable for a class of email systems.

This is a new system SARE is experimenting with, and so file names may change as we develop improvements to this system, but the ideas should remain basically the same.

Each multi-file rule set has at least four files, and several have more. These files are:

- File 0 contains those rules which hit spam and only spam (no ham), and which hit significant spam (more than 10 spam among several corpora). Those are rules which we feel confident every system will benefit from. (Warning: email administrators may still need to review topic-specific rules within this file to verify they are appropriate for your system. Example: systems which support pharmacies or druggists may wish to exclude drug-specific rules.)
- File 1 contains rules which either hit fewer than 10 spam, or which also hit ham. The primary qualification for the latter is that our S/O (total spam count / total email count) must be 0.900 or higher. This file is also probably good for most sites, but these rules are not as conservative, not as safe, as those in file 0.
- File 2 contains rules which should hit spam and only spam, which SARE considers to be conservative and safe to use, but recent mass-checks had no matches for these rules. Many of these are obfuscation rules, designed to hit only obfuscated words or phrases (vi@gra but not viagra). Aggressive systems with more than sufficient resources may benefit from these rules eventually (some will hit spam some day), but those systems with tight resources should definitely stay away from these.
- File 3 contains rules which hit lots of ham. They are scored low, and might help push otherwise questionable spam over the SA threshold. Very conservative systems, and those that are tight on resources, should probably avoid these files. (Note that as the behavior of spam changes, rules may be dropped from this file simply because they stop hitting enough spam to make them worth while.)
- Any file named *_eng.cf contains rules which apply to the English language, but which can readily cause false positives in other languages. Rules which attempt to identify spam by the number of consecutive consonants is one example of this type rule (the German language uses more consecutive consonants than English does). Systems in countries where English is not the native language, and systems whose users receive emails in other languages, should avoid this rule set file.
- Any file named *_arc.cf contains archived rules, rules which used to hit spam, but no longer do, and which do not qualify for file # 2 above. These rules are kept here, and tested regularly by SARE, so the rules can be reactivated (moved to other files) if/when they again start hitting spam. Aggressive systems with plenty of resources may wish to include this file so they gain the benefit of these rules without waiting for SARE's (monthly?) test cycle.
- Other files are version-specific.
 - A file named *_v25.cf contains arithmetic meta rules, a functionality that became available in version 2.50 of [SpamAssassin](#), and which does not work in any 2.4x or earlier version.
 - A file named *_v24.cf contains non-arithmetic meta rules which emulate the arithmetic rules mentioned above. Systems running SA 2.4x or earlier versions should use this file instead of the *_v25.cf file.
 - A file named *_v30.cf contains functionality which works only with SA version 3.0.0 or higher. Systems running 2.xx versions of [SpamAssassin](#) should avoid this file.
 - A file named *_x30.cf contains rules which duplicate distribution rules found in version 3.0.0. Systems running version 3.0.0 should avoid this file.Other files may be named similarly. If their use is not obvious by the name, it should also be documented within the file.